Alma Mater Studiorum - Università di Bologna
Facoltà di Scienze Matematiche, Fisiche e Naturali
Dipartimento di Astronomia

DOTTORATO DI RICERCA IN ASTRONOMIA
CICLO XXIV

TESTING FUTURE WEAK LENSING SURVEYS THROUGH SIMULATIONS OF OBSERVATIONS

Tesi di Dottorato di
FABIO BELLAGAMBA

Coordinatore Dottorato
Prof. **LAURO MOSCARDINI**

Relatori
Prof. **LAURO MOSCARDINI**
Dott. **MASSIMO MENEGHETTI**

Esame finale anno 2012

SCUOLA DI DOTTORATO IN SCIENZE MATEMATICHE, FISICHE E ASTRONOMICHE
Settore concorsuale: 02/C1 - Astronomia, Astrofisica, Fisica della Terra e dei Pianeti
Settore scientifico-disciplinare: FIS/05 - Astronomia e Astrofisica

Contents

Introduction

In the last decades the scientific community was able to build a cosmological standard model which explains coherently many observational data. This model, based on the gravitational theory condensed in Einstein's General Relativity, is able to describe properly both the measured expansion of the Universe and the formation of cosmic structures we observe. According to the standard model, the baryonic matter that forms planets, stars and galaxies represent only a tiny fraction ($\sim 4\%$) of the total matter/energy content of the Universe. The major components are the so-called Dark Matter, which drives the gravitational collapse and accounts for $\sim 23\%$ of the matter/energy content, and the Dark Energy, which gives rise to the accelerated expansion of the Universe and is nowadays dominant, containing $\sim 73\%$ of the energy of the Universe. Unfortunately, our knowledge of the nature of these two components is very limited. We detect Dark Matter through the gravitational force it produces, but its electromagnetic interactions must be very weak as we are not able to observe it emitting or absorbing photons. Dark Energy is even more subtle: we do not even know if it exists as a separate field or it is the consequence of some unknown physics.

In order to gain new insights on these mysterious components of the Universe, and to improve the very basic description we are able to offer, a new experimental effort is needed. From this point of view, a very useful cosmological tool is cosmic shear, i.e. the analysis of the small distortions in the shapes of observed galaxies produced by the intervening large-scale structure. The strength of this observable lies in the fact that few physical assumptions are needed to extract cosmological information from the data. In fact, the lensing effect is due to the total matter distribution along the line of sight, irrespectively of its nature, temperature, pressure, chemistry and so on. From the statistical analysis of observed shapes of galaxies it is possible to draw

very precise measurements of cosmological parameters, such as the equation of state of Dark Energy. Moreover, with large data-sets of galaxies, such as those obtained from weak lensing observations, other important cosmological probes can be derived, such as the abundance of galaxy clusters as a function of mass and redshift. Clusters of galaxies grew from the highest peaks in the primordial density fluctuations and are the largest objects which are gravitationally bound. Their distribution thus describes both the original density field and the history of cosmic growth.

To unfold all the cosmological potential of these observables, many dangerous systematics must be put under control. In particular, the measure of cosmic shear needs a large angular density of observed galaxies to kill the statistical uncertainty due to the distribution of intrinsic shapes and a very good redshift measurement to slice the sample and observe the growth of cosmic structures as a function of redshift. The employment of clusters of galaxies as cosmological tools depend on the ability to detect complete and pure samples from the data, and to obtain a robust measurement of their mass. Simulations of observations, that reproduce in a realistic way many sources of noise present in the real data, are thus a very useful way to test performances of instruments, surveys and analysis tools. Submitting mock observations to programs that are usually run on real data, it is possible to test the performances of future observations, and the precision of output data that will be used for cosmology.

In this work we show our technique to perform simulations of observations and we apply this tool to a future space-based mission named Euclid. Euclid is an ESA-approved satellite that aims to measure with unprecedented precision the expansion of the Universe and the formation of structures through two main independent observables: cosmic shear and galaxy clustering. In particular we will focus on the precision of photometric redshift measurements and on the amount of clusters that will be detected in Euclid data. The outline of the Thesis is as follows:

- In Chapter 1 we will overview the description of the history of the Universe which makes up the cosmological standard model. We will describe first the evolution of the smooth background, focusing on the accelerating expansion which shows us the effect of Dark Energy, and then on the formation of cosmic structures and on how the biggest of

them can be used as tracers of cosmological parameters.

- In Chapter 2 we will introduce gravitational lensing in general and then focus on cosmic shear, both from the theoretical and from the observational point of view. In the last Section, we introduce the main characteristics and expected performances of the Euclid mission.

- In Chapter 3 we will describe our code to perform realistic simulations of observations and we will show an application to the estimate of cluster masses from lensing data.

- Chapter 4 is devoted to the discussion of cluster detection from optical data. In particular, we will describe the construction of an optimal filter for cluster detection from galaxy distribution and we will show its usage on real data from the COSMOS survey.

- In Chapter 5 we will apply the techniques described in the previous Chapters to the Euclid survey. We will first use mock images to estimate the precision of photometric redshift measurements, and then we will make use of this knowledge to assess the ability of detecting galaxy clusters from photometric catalogues.

- In the Conclusions we will summarise the research work presented in this Thesis and present our main results.

Chapter 1

Cosmology

1.1 Evolution of background

1.1.1 Robertson-Walker metric

The global evolution of the Universe is governed by gravity, which is described by Einstein's General Relativity. This theory connects the geometry of the four-dimensional space-time to the distribution of energy, including that in the form of matter or radiation. To describe the global dynamics of the Universe, we need to make some assumptions about its symmetries. In particular, we postulate the so-called Cosmological Principle, which states that the Universe is homogeneous and isotropic. This means that the three-dimensional space geometry must be invariant under rotations or translations. This assumption is supported by the observation of the Cosmic Microwave Background, the imprint of the conditions of the Universe at its early stage, whose inhomogeneities are at the 10^{-5} level. At the present age, the existence of stars and galaxies means that the homogeneity is valid only for certain scales, namely those larger than the biggest structures. However, this does not affect our considerations as long as these scales are much smaller of the cosmological distances between these structures. In fact, we will first describe the evolution of an ideal smoothed Universe and then focus on how residual inhomogeneities grew to form cosmic structures.

There are three geometrical surfaces that fulfil the aforementioned assumptions, with the following line elements:

- hypersphere
$$dl^2 = \frac{dr^2}{1-r^2} + r^2(d\theta^2 + \theta^2 d\phi^2) \tag{1.1}$$

- hyperplane
$$dl^2 = dr^2 + r^2(d\theta^2 + \theta^2 d\phi^2) \tag{1.2}$$

- hyperboloid
$$dl^2 = \frac{dr^2}{1+r^2} + r^2(d\theta^2 + \theta^2 d\phi^2)\,. \tag{1.3}$$

In the previous definitions, r represents the distance coordinate, while θ and ϕ are the angular coordinates. The different line elements can be summarised in the form
$$dl^2 = \frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \theta^2 d\phi^2)\,, \tag{1.4}$$
where the constant K is 0 in case of a flat geometry, +1 in case of a closed geometry, and -1 in case of an open one.

The Cosmological Principle allows us to define the so-called *cosmic time* t, that is the proper time of a particle travelling perpendicular to the homogeneous three-dimensional space. Given the symmetry of the Universe, its evolution is described by a single scalar function $a(t)$, that modulates the distance scale as a function of time. The complete space-time line element is thus
$$ds^2 = (cdt)^2 - a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right], \tag{1.5}$$
the so-called *Robertson-Walker* metric. The geometry of the space-time, generally contained in the tensor G, is thus described by the constant K and the scalar function $a(t)$ only. They are linked to the energy content of the Universe through the Einstein equation,
$$G_{\alpha\beta} = \frac{8\pi G}{c^2} T_{\alpha\beta} + \Lambda g_{\alpha\beta}\,, \tag{1.6}$$
where T is the energy-momentum tensor and the last term, proportional to the metric tensor g, was introduced by Einstein to allow the existence of a static Universe. The content of the Universe is generally described as a perfect fluid, where the only non-zero terms of T are the diagonal ones, explicitly $(\rho c^2, p, p, p)$.

1.1.2 Redshift

As the scale factor $a(t)$ varies with time, the wavelength of the photons changes during their path from the place where they are emitted to the observer. In particular, in an expanding phase of the Universe, as the one we are experiencing, the so-called *cosmological redshift* takes place. If the coordinates of the source and of the observer do not change as a function of time, leaving apart the effect of the expansion of the Universe, their (comoving) distance is

$$d_c = \int_0^r \frac{dr'}{(1-Kr')^{1/2}} = \int_{t_e}^{t_o} \frac{cdt}{a(t)} = \int_{t_e+\delta t_e}^{t_o+\delta t_o} \frac{cdt}{a(t)}\,. \tag{1.7}$$

Considering two photons that were emitted in times separated by a little interval δt_e this implies

$$\frac{\delta t_e}{a(t_e)} = \frac{\delta t_o}{a(t_o)}\,, \tag{1.8}$$

where δt_o is the (little) interval between their arrival times. Thus, as the emitted frequency ν_e is inversely proportional to the emitting time, we can write

$$\frac{\delta t_o}{\delta t_e} = \frac{\nu_e}{\nu_o} = \frac{\lambda_o}{\lambda_e}\,. \tag{1.9}$$

The redshift z is defined as

$$z \equiv \frac{\lambda_o - \lambda_e}{\lambda_e}\,. \tag{1.10}$$

From Eqs. (1.8) e (1.9) we can get the relation between the redshift and scale factor:

$$1 + z = \frac{a(t_o)}{a(t_e)}\,. \tag{1.11}$$

The redshift of an astronomical source can then be interpreted as a measure of the expansion suffered by the Universe between the times in which the photons were emitted and observed.

1.1.3 Friedmann equations

Given the assumptions we made on the geometry and the energy content of the Universe, equation (1.6) can be converted in the so-called Friedmann equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3} \tag{1.12}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{Kc^2}{a^2} + \frac{\Lambda}{3}\,, \tag{1.13}$$

where dots represent derivatives with respect of time.

In many interesting cases, the relation between pressure and density of a perfect fluid can be expressed as

$$p = w\rho c^2\,. \tag{1.14}$$

In particular, we have $w \sim 0$ for non-relativistic matter and $w = 1/3$ for radiation. In the rest of this Section we assume that the free parameter Λ, called *cosmological constant*, is null. We will come back to this point in Sect. 1.1.7. In case of a flat geometry, the Friedmann equations can then be solved to write

$$a(t) = a_0(t/t_0)^{\frac{2}{3}(1+w)}\,. \tag{1.15}$$

Thus, the expansion of the Universe proceeds as

$$a(t) = a_0(t/t_0)^{2/3} \tag{1.16}$$

in case the fluid is made by matter, and

$$a(t) = a_0(t/t_0)^{1/2} \tag{1.17}$$

if it is made by radiation. It is worthful to notice that in both cases the expansion of the Universe is decelerated by the gravitational attraction acting on the fluid itself. The derivative $\dot{a}$ goes as $(t/t_0)^{-\frac{1}{3}(1+w)}$ and thus the velocity of the expansion remains always positive but tends to zero when the time goes to infinity. In case the geometry is not flat, Eqs. 1.16 and 1.17 are not valid anymore. In particular, a closed universe will reach a maximum value of a, where $\dot{a}$ is equal to zero, after which the expansion stops and the universe collapses back. On the other hand, an open universe will continue its expansion indefinitely, with a speed $\dot{a}$ that remains positive and does not tend to zero, in opposition to the flat case.

A general feature of these models is that for small values of t, a tends to zero, indicating the existence of a period when the dimensions of the Universe were extremely small, from which the so-called *Big Bang* started. The validity of this description ends when the size of the Universe becomes so small that the Heisenberg principle comes into play on cosmological scales.

This is the *Planck era*, that lasted around 10^{-43} seconds, when the quantum fluctuations were active on the scale of the cosmological horizon, and therefore are not negligible, neither in first approximation.

1.1.4 Critical density

The actual geometry of the space-time depends on the energy content of the Universe. In fact, Eq. 1.13 can be rewritten as

$$H^2 = \frac{8\pi G\rho}{3} - \frac{Kc^2}{a^2}\,, \tag{1.18}$$

where we substituted the *Hubble constant* H to $\dot{a}/a$. At the present time t_0, the relation between the geometry parameter K and the energy density is

$$\rho_0 = \frac{3}{8\pi G}\left(H_0^2 + \frac{Kc^2}{a^2}\right), \tag{1.19}$$

where we indicated with the subscript 0 the quantities evaluated at t_0. We can then define the *critical density* ρ_{cr} as the density the Universe would have if the geometry was flat:

$$\rho_{cr,0} = \frac{3H_0^2}{8\pi G}\,. \tag{1.20}$$

If the present density of the Universe is bigger than $\rho_{cr,0}$, we have $K > 0$ and the geometry is closed; if it is smaller than $\rho_{cr,0}$, K is negative and the geometry is open. (K can always be rescaled to 1 or -1, changing accordingly a and r, without affecting any physical quantity.) It is usual to rephrase this in terms of the density parameter Ω_0, defined as

$$\Omega_0 \equiv \frac{\rho_0}{\rho_{cr,0}}\,. \tag{1.21}$$

Then, the geometry of the Universe will change if Ω_0 is bigger, equal to, or smaller than 1. It is important to note that to this end it is not important what the Universe is made of; if different components are present, the geometry of the Universe will be determined by the sum of their densities. Thus, Eq. 1.20 can be used as an absolute scale to measure the amount of the each ingredient. For example, we define Ω_m as the contribution of the matter to the density parameter.

As the energy of the radiation in contained mainly in the (*Cosmic Microwave Background*, CMB), that has today a temperature $T_{CMB} \simeq 2.73K$, its contribution to the critical density Ω_0 is

$$\Omega_{CMB,0} = 2.4 \times 10^{-5} h^{-2} \,. \tag{1.22}$$

Relativistic matter, as neutrinos, has the same equation of state of radiation and thus can be summed up in this perspective, obtaining a total contribution

$$\Omega_{R,0} = 3.2 \times 10^{-5} h^{-2} \,. \tag{1.23}$$

The energy density of baryonic matter is estimated via primordial nucleosynthesis and CMB mainly. The results converge on

$$\Omega_{B,0} \sim 0.04. \tag{1.24}$$

The existence of another type of matter, called *dark matter* was first deduced studying the dynamics of clusters and galaxies, and it has then been confirmed by gravitational lensing and structure formation studies. This matter behaves like the ordinary one with respect to gravity, but it must have very weak electromagnetic interactions, as we cannot observe it in any wavelength range. Its density is estimated to be around 5 times the baryonic matter one. The total contribution of matter to the critical density is thus

$$\Omega_{M,0} \sim 0.3 \,. \tag{1.25}$$

The Λ term in Friedmann equations can be interpreted as the result of a negative pressure fluid, called *dark energy*. In this case, its contribution would be the one missing to achieve the observed flatness of the Universe, namely

$$\Omega_{\Lambda,0} \simeq 0.7 \,. \tag{1.26}$$

1.1.5 Cosmological distances

In a curved space-time, the concept of distance is no more straightforward as in the Euclidean three-dimensional space. Different definition of distances can be used to different purposes.

The *proper distance* d_P is the distance measured by a chain of observers connecting two points at time t. From Eq. 1.4 one gets

$$d_P = \int_0^r \frac{a dr'}{(1 - Kr')^{1/2}} = a f(r)\,, \tag{1.27}$$

where a is the scale factor at time t, and $f(r)$ is

$$f(r) = \begin{cases} arcsin\, r & (K = 1) \\ r & (K = 0) \\ arcsinh\, r & (K = -1). \end{cases} \tag{1.28}$$

The proper distance at the actual time t_0 is the *comoving distance* d_c, which we already introduced in Sect. 1.1.2.

The *luminosity distance* d_L is defined in such a way that the flux F received by a source with luminosity L follows the Euclidean law

$$F = \frac{L}{4\pi d_L^2}\,, \tag{1.29}$$

or equivalently

$$d_L = \left(\frac{L}{4\pi F}\right)^{1/2}. \tag{1.30}$$

In their path the photons suffer a loss of energy proportional to a, and a dilatation of time again proportional to a. The received flux F is a^2 times weaker than it would be in an Euclidean space. Thus, to keep Eq. 1.30 valid, the luminosity distance must be

$$d_L = r \times a^{-1}\,. \tag{1.31}$$

Analogously, the *angular diameter distance* is the one that conserves the geometrical law that connects the physical dimension d_P of an object to the subtended angle $\Delta\theta$ when seen from an observer.

$$\Delta\theta = \frac{d_P}{d_A}\,. \tag{1.32}$$

From Eq. 1.5, one derives $d_P = ar\Delta\theta$ and thus

$$d_A = ar = d_L a^2\,. \tag{1.33}$$

1.1.6 Expansion history of the Universe

We saw in Sect. 1.1.3 the solutions to the Friedmann equations in case of single-component universes. As we saw in the Sect. 1.1.4, our Universe is made of different ingredients. In this case, the Eq. 1.13 generalises to

$$H^2(t) = H_0^2\left[\Omega_{R,0}\left(\frac{a}{a_0}\right)^{-4} + \Omega_{M,0}\left(\frac{a}{a_0}\right)^{-3} + \Omega_{\Lambda,0} - \frac{Kc^2}{H_0^2}\left(\frac{a}{a_0}\right)^{-2}\right]. \quad (1.34)$$

At very early epochs, that correspond to very small values of a, Ω_R dominates the second term of the equation, and thus the expansion. As we have shown in Sect. 1.1.3, the behaviour of the scale factor as a function of time in this case is $a(t) \propto a^{1/2}$. It is important to underline that the geometry of the Universe at early times must be very close to flatness. In fact, one can derive from the Friedmann equations the following relation:

$$\Omega^{-1}(t) - 1 = \frac{\Omega_0^{-1} - 1}{a^{-(1+3w)}}\,. \quad (1.35)$$

For smaller and smaller values of a, $\Omega(t)$ becomes closer and closer to unity. Moreover Eq. 1.35 shows that, if $\Omega_0 > 1$ ($\Omega_0 < 1$, $\Omega_0 = 1$) at time t_0, the same relation must hold at any time t.

We define the *equivalence time* as the moment matter reached the same energy density of radiation, and then became dominant:

$$a_{eq} = \frac{\Omega_M}{\Omega_R}\,. \quad (1.36)$$

After a_{eq} we can approximate the expansion of the Universe to be matter-dominated and thus follow $a(t) \propto t^{2/3}$. At an even later time, the Λ component becomes dominant. We can see its effect on the expansion history of the Universe, by writing

$$K = \left(\frac{H_0 a_0}{c}\right)^2 (\Omega_{M,0} + \Omega_{\Lambda,0} - 1)\,, \quad (1.37)$$

where we considered that $\Omega_{R,0} \ll \Omega_{M,0}$. Thus we have

$$H^2(t) = H_0^2\left[\Omega_{R,0}\left(\frac{a}{a_0}\right)^{-4} + \Omega_{M,0}\left(\frac{a}{a_0}\right)^{-3} + \Omega_{\Lambda,0} - (1 - \Omega_{M,0} - \Omega_{\Lambda,0})\left(\frac{a}{a_0}\right)^{-2}\right]. \quad (1.38)$$

To understand the evolution of the scale factor as a function of time, it is useful to introduce the *deceleration parameter* q, defined as

$$q(t) \equiv -\frac{\ddot{a}(t)a(t)}{\dot{a}^2(t)}\,, \tag{1.39}$$

which is positive when the expansion of the Universe is slowing down. Using Eqs (1.12), (1.13) and (1.37), the actual value of the deceleration parameter can be written as a function of $\Omega_{M,0}$ and $\Omega_{\Lambda,0}$. One gets

$$q_0 = \frac{\Omega_{M,0}}{2} - \Omega_{\Lambda,0}\,. \tag{1.40}$$

In presence of a dark energy component (or a cosmological constant), which we will discuss in the next Section, if $\Omega_{\Lambda,0} > \frac{\Omega_{M,0}}{2}$, q_0 becomes negative and we have an accelerating expansion of the Universe, which cannot be obtained just with radiation and matter. In particular, extrapolating for large values of t, for which we can totally neglect the other components, the expansion proceeds as

$$a(t) \propto \exp\left[\left(\frac{1}{3}\Lambda\right)^{1/2} ct\right], \tag{1.41}$$

in a so-called *De Sitter universe*.

1.1.7 Accelerated expansion and Dark Energy

Two teams working independently in the mid-to-late 1990s, the Supernova Cosmology Project and the High-z Supernova Search, found that distant supernovae are 0.25 mag dimmer than they would be in a decelerating universe (see Fig. 1.1), indicating that the expansion has been speeding up for the past 5 Gyr (Riess et al., 1998; Perlmutter et al., 1999). This result alone, in the framework of General Relativity, proves the presence of a component of the universe with $w < 1/3$. Since then, many different cosmological probes (CMB spectrum, cosmic shear, baryon acoustic oscillations) converged on a cosmological model where a dark energy component is dominant at the present epoch, $\Omega_\Lambda \simeq 0.7$. The equation of state of this component is compatible with $w = -1$, the same that a cosmological constant term in Friedmann equations would create (see Fig. 1.2).

There are many different proposals about the nature of dark energy, ranging from the quantum vacuum to a new, ultralight scalar field. Alternatively,

cosmic acceleration may arise from new gravitational physics, perhaps involving extra spatial dimensions. For dark energy, the equation-of-state parameter w provides a useful phenomenological description (Turner & White, 1997). As it is the ratio of pressure to energy density, it is also closely connected to the underlying physics. Its evolution with redshift is usually parametrised through

$$w(a) = w_0 + (1-a)w_a = w_0 + \frac{z}{1+z}w_a\,. \tag{1.42}$$

Although the existence of cosmic acceleration is now well established, and the dark energy density is determined at the precision of a few percent (see Fig. 1.3), trying to catch its nature by measuring the evolution of its equation of state is much more challenging. Apart from supernovae observations, the other three main probes of dark energy nature are clusters of galaxies, BAO and cosmic shear. The amount of clusters of galaxies and the strength of cosmic shear signal are sensitive both to the effect of dark energy on the expansion of the Universe and to its consequences on the growth of cosmic structures. We will come back to these probes later, after having introduced the theory of cosmic structure formation in the next Section.

1.2 Formation of cosmic structures

1.2.1 Jeans theory

The standard model of structure formation is based on the growth of little fluctuations in the density distribution through gravitational collapse. The main quantity in the following discussion is the *density contrast* δ, defined as

$$\delta(\vec{x}, a) \equiv \frac{\rho(\vec{x}, a) - \bar{\rho}(a)}{\bar{\rho}(a)}\,, \tag{1.43}$$

where $\bar{\rho}(a)$ is the mean density of the considered component. As long as the perturbation is small ($\delta \ll 1$) we can use the linear theory of perturbation; to analyse the evolution of the perturbation from the moment in which $\delta \simeq 1$ we need different analytic approximations or numerical simulations. In the linear regime, if we Fourier transform the density field, each fluctuation with wavelength $k = 2\pi/\lambda$ evolves independently following

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} + (v_s^2 k^2 - 4\pi G\rho)\delta = 0\,, \tag{1.44}$$

where $v_s = (\partial p/\partial \rho)^{1/2}$ is the speed of sound. Thus, a perturbation can grow when the force of gravity that acts to make it collapse prevails over the expansion of the Universe and on the pressure which tends to bring it back to equilibrium. This happens usually for fluctuations above a certain scale, called *Jeans scale* λ_J, which depends on the kind of fluid the perturbation is made of and on the background mean conditions. When the component that tends to collapse is the dominant one, we have

$$\lambda_J \simeq v_s \left(\frac{\pi}{G\rho} \right)^{1/2} . \tag{1.45}$$

Another important scale is the *horizon scale* d_H, which indicates the size of the regions of the Universe which are causally connected at a given cosmic time. Using as a temporal scale the inverse of the Hubble constant, the photons at a certain time t travelled a comoving distance

$$d_H(a) = \frac{c}{H(a)a} . \tag{1.46}$$

Beyond this scale, the Jeans theory is no more correct and we need a purely relativistic treatment, where the fluctuations in the metric tensor $g_{\alpha\beta}$ are considered. The relativistic and non-relativistic treatment, in the linear regime, show that the matter perturbations grow as

$$\delta(a) \propto \begin{cases} a^2 & \text{before } a_{eq} \\ a & \text{after } a_{eq}. \end{cases} \tag{1.47}$$

So the behaviour of the perturbations changes depending on the component (matter or radiation) that dominates the Universe. Eq. 1.47 is valid exactly in a flat universe with $\Omega_{M,0} = 1$ and $\Omega_{\Lambda,0} = 0$. In the general case, we have

$$\delta(a) = \delta_0 a \frac{g'(a)}{g'(a_0)} = \delta_0 a g(a) , \tag{1.48}$$

where δ_0 is the density contrast at the present epoch and the *growth function* g' can be approximated as (see Carroll et al., 1992)

$$g'(a; \Omega_{M,0}; \Omega_{\Lambda,0}) = \frac{5}{2}\Omega_{M,0} \left[\Omega_M^{4/7}(a) - \Omega_\Lambda(a) + \left(1 + \frac{\Omega_M(a)}{2} \right) \left(1 + \frac{\Omega_\Lambda(a)}{70} \right) \right]^{-1} . \tag{1.49}$$

The behaviour of the growth function for different cosmological parameters can be seen in Fig. 1.4.

The observations of the Cosmic Microwave Background (CMB) reveal temperature density fluctuations of the order of 10^{-5}. The matter-radiation decoupling happened at $z \sim 1000$, at a time $a_{dec} \gg a_{eq}$. Applying Eq. 1.47 the density fluctuations should be of order of 10^{-2} at the present epoch. This goes against the observation of collapsed structures, with $\delta \gg 1$. This apparent contradiction is solved invoking dark matter, which we already introduced in Sect. 1.1.4. While radiation and matter have been linked through electromagnetic processes until decoupling, dark non-collisional matter decoupled much earlier and started to collapse on density peaks. These fluctuations, which did not leave any imprint on CMB, are the ones which gave rise to structures we observe nowadays. After decoupling, baryonic matter has been free to fall in the already formed dark matter potential wells, thus growing much faster than in the linear prescription.

1.2.2 Power spectrum

In the standard cosmological model, the primordial inhomogeneities derive from quantum fluctuations originated in an early phase of exponential expansion, called *inflation*. Many models suggest that these fluctuations should follow a Gaussian statistic. In this case, their distribution is completely determined by the *power spectrum* $P_\delta(k)$, defined by

$$\langle \hat{\delta}(\vec{k}) \hat{\delta}^*(\vec{k'}) \rangle = (2\pi)^3 \delta_D(\vec{k} - \vec{k'}) P_\delta(k)\,, \tag{1.50}$$

where $\hat{\delta}(\vec{k})$ is the Fourier transform of δ. It is generally assumed that the primordial power spectrum is described by a power law, $P_i(k) \propto k^n$. Given that before a_{eq} the density contrast grows as $\delta \propto a^2$, the spectrum grows as $P(k) \propto a^4$. A perturbation with comoving scale λ enters in the horizon when $\lambda = d_H(a)$. For all scales $\lambda < d_H(a_{eq})$, or equivalently $k > k_H = d_H^{-1}(a_{eq})$, the spectrum at the moment they enter the horizon a_{enter} is thus

$$P_{enter}(k) \propto a^4 P_i(k) \propto k^{-4} P_i(k)\,. \tag{1.51}$$

It is generally accepted that the total power of the fluctuations when they enter the horizon should be *scale-invariant*, that is $k^3 P(k) = cost$. This

brings $P_i(k) \propto k$. This spectrum is called *Harrison-Zel'dovich spectrum*. If $k > k_H$, the perturbation enters the horizon when the radiation is still dominant ($a_{enter} < a_{eq}$). In this case its growth stops, because the rapid expansion produced by radiation prevents the collapse. We can see this also by comparing the characteristic time scale of the expansion of the Universe, t_{exp}, with the characteristic time scale of the gravitational collapse:

$$t_{exp} \sim (G\rho_R)^{1/2} \ll (G\rho_M)^{1/2} \sim t_{coll}\,. \tag{1.52}$$

Outside the horizon, on scales that the photons cannot have travelled yet, the growth proceeds as $\delta \propto a^2$. Thus the dimension of the horizon at the equivalence $d_H(a_{eq})$ is an important scale for the evolution of the perturbations: the ones with $k > k_H$ suffer the so-called *stagnation*, and are decremented by a factor $(a_{enter}/a_{eq})^2$.

Going back to the power spectrum, this translates in

$$P_\delta(k) \propto \begin{cases} k^n & \text{for } k \ll k_0 \\ k^{n-3} & \text{for } k \gg k_0 \end{cases} \tag{1.53}$$

at equivalence. In particular, for a Harrison-Zeldovich spectrum this becomes

$$P_\delta(k) \propto \begin{cases} k & \text{for } k \ll k_0 \\ k^{-3} & \text{for } k \gg k_0 \end{cases} \tag{1.54}$$

From then on, the dark matter perturbation can grow around density peaks following Eq. 1.47 on every scale, because the Jeans scale for dark matter perturbations, once they are decoupled from radiation, is negligible from the cosmological point of view.

1.2.3 Nonlinear regime

Once the density perturbations enter the nonlinear regime, their evolution cannot be described with the equations we introduced above. In particular, the different modes denoted by their scale k do not evolve independently anymore. To make an analytic description of the collapse one must make some strong assumptions on the initial conditions of the system. The simplest case is that of a spherically symmetric density perturbation that collapses

along the radial direction. In absence of a cosmological constant, the radius R of a mass shell will obey the equation

$$\frac{d^2R}{dt^2} = -\frac{GM}{R^2}\,, \tag{1.55}$$

where M is the mass inside the shell. Integrating this equation, one obtains the energy budget of the system

$$\frac{1}{2}\left(\frac{dR}{dt}\right)^2 - \frac{GM}{R} = E\,. \tag{1.56}$$

If $E < 0$, the shell collapses. In particular, the maximum expansion of the system, with $R = R_{max}$ at $t = t_{max}$, is obtained when the density of the perturbation is $\rho(t_{max}) = (3\pi/4)^2\bar{\rho}(t_{max}) \simeq 5.5\bar{\rho}(t_{max})$, if $\bar{\rho}$ is the average density of the background. Then R starts to decrease and reaches the singularity at $R = 0$ at the collapse time $t_c = 2t_{max}$, when the density ideally reaches infinity at the center.

In fact, when the density is high, small departures from spherical symmetry will result in the formation of shocks and important pressure gradients. Heating of the material will occur due to the dissipation of shocks which converts the kinetic energy of collapse into heat. The final result will therefore be an equilibrium state which is not a singular point but some extended configuration with radius R_{vir} and mass M. This happens when the system reaches the virial equilibrium. From the virial theorem the total energy of the perturbation is

$$E = -\frac{1}{2}\frac{3GM^2}{5R_{vir}}\,. \tag{1.57}$$

If we equalize this value with that of the perturbation at time t_{max},

$$E = -\frac{3GM^2}{5R_{max}}\,. \tag{1.58}$$

we see that the virial radius R_{vir} is half R_{max}. Therefore, the density of the perturbation in the equilibrium state is $\rho(t_{vir}) = 8\rho(t_{max})$. Numerical simulations of the collapse allow an estimate of the time taken to reach equilibrium: one finds that $t_{vir} \simeq 3t_{max}$. The perturbation density in terms of the critical density at t_c is

$$\Delta_c \equiv \frac{\rho(t_c)}{\rho_{cr}(t_c)} = 18\pi^2 \simeq 178\,. \tag{1.59}$$

An incorrect extrapolation of the linear results would give a density contrast $\delta_c(t_c) = 1.68$, which corresponds to a value of 2.68 for the ratio of the densities, instead of the value of Eq. 1.59.

The real effects of the nonlinear regime on the power spectrum can be calculated only through numerical simulations. The scales that are more affected by the nonlinear evolution are the smallest ones, as can be seen in Fig. 1.5.

The previous sections give the theoretical expectations for the shape and growth rate of the density perturbation spectrum but do not specify its normalization. Because inflationary theories do not make firm predictions about the amplitude of the primordial power spectrum, the normalization of $P(k)$ must be determined observationally. For example, measurements of the present-day mass distribution of the Universe indicate that $\delta_M/M \simeq 1$ within comoving spheres of radius 8 h^{-1} Mpc, as suggested by early galaxy surveys showing that the variance in galaxy counts was of order unity on this length scale (Davis & Peebles, 1983). This feature of the Universe is the motivation for expressing the power-spectrum normalization in terms of the quantity σ_8, where

$$\sigma_8^2 = \frac{1}{(2\pi)^3} \int P(k)|W_k|^2 d^3k \tag{1.60}$$

is the variance defined with respect to a top-hat window function $W(r)$ having a constant value inside a comoving radius of 8 h^{-1} Mpc and vanishing outside this radius. When using this formula, one must keep in mind that $P(k)$ refers to the power spectrum of linear perturbations evolved to $z = 0$ according to the growth function $\delta(z)$, which is valid only for small perturbations.

1.2.4 Mass function

The *mass function* $n(M)$ of cosmic structures is defined by the relation

$$dN = n(M)dM \tag{1.61}$$

which gives the number of structures per unit volume with mass between M and $M + dM$. Press & Schechter (1974) proposed a model to calculate analytically the mass function using some knowledge derived from the spherical collapse approximation. This formalism first smooths the density field $\delta(\vec{x})$,

by filtering it on a spatial scale R, defining a field δ_M,

$$\delta_M = \delta(\vec{x}; R)\,, \tag{1.62}$$

where the mass M is related to the scale R through the mean density of the Universe. If the density field possesses Gaussian statistics, the distribution of the fluctuations is given by

$$P(\delta_M)d\delta_M = \frac{1}{(2\pi\sigma_M^2)^{1/2}} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2}\right) d\delta_M\,. \tag{1.63}$$

The probability that the smoothed density field exceeds at some point a threshold value δ_T is by construction

$$P_{>\delta_T}(M) = \int_{\delta_T}^{\infty} P(\delta_M)d\delta_M\,. \tag{1.64}$$

It is interesting to set this threshold to $\delta_c \sim 1.68$, that is the density contrast derived from linear theory of structures which have collapsed and reached their maximum density. If we want to count the objects with mass in the range between M and $M + dM$, we must then calculate

$$n(M)MdM = 2\rho_{max}[P_{>\delta_c}(M) - P_{>\delta_c}(M + dM)] = 2\rho_{max}\left|\frac{dP_{>\delta_c}}{d\sigma_M}\right|\left|\frac{d\sigma_M}{dM}\right| dM\,, \tag{1.65}$$

where the factor 2 accounts for accretion from underdense regions, which represent half of the field, but would not otherwise be taken into account. If the RMS mass fluctuation can be expressed by a power law,

$$\sigma_M = \left(\frac{M}{M_0}\right)^{-\alpha}, \tag{1.66}$$

one obtains the following expression for the mass function,

$$n(M) = \frac{2}{\sqrt{\pi}}\frac{\rho_{max}\alpha}{M_\star^2}\left(\frac{M}{M_\star}\right)^{\alpha-2} \exp\left[-\left(\frac{M}{M_\star}\right)^{2\alpha}\right]. \tag{1.67}$$

Thus, the mass function has a power-law behaviour with an exponential cut-off for masses bigger than

$$M_\star = \left(\frac{2}{\delta_c^2}\right)^{1/2\alpha} M_0\,. \tag{1.68}$$

An improvement of this calculation has been done by Sheth & Tormen (1999), considering the ellipsoidal collapse instead of the spherical one. The major modification is an increase of objects at the high mass tail, in better agreement with the results of N-body simulations.

1.2.5 Galaxy clusters

Clusters of galaxies are the largest virialised structures of the Universe. They trace the high-mass tail of the mass function, and therefore their abundance as a function of mass and redshift can be used to constrain the cosmological parameters. This can be seen if we write the Sheth and Tormen result in a differential expression,

$$\frac{dn_M}{d\ln\sigma^{-1}} = A_s\sqrt{\frac{2a_s}{\pi}}\left[1+\left(\frac{\sigma^2}{a_s\delta_c^2}\right)^{p_s}\right]\frac{\Omega_M\rho_{cr,0}}{M}\frac{\delta_c}{\sigma}\exp\left[-\frac{a_s\delta_c^2}{2\sigma^2}\right], \qquad (1.69)$$

with $A_s = 0.3222$, $a_s = 0.707$, and $p_s = 0.3$. The variance σ depends on mass M and redshift z through the linear evolution of the primordial spectrum

$$\sigma^2(M,z) = \frac{g^2(z)}{(2\pi)^3}\int P(k)|W_k(M)|^2 d^3k\,, \qquad (1.70)$$

where $W_k(M)$ is the Fourier-space representation of a top-hat window function that encloses mass M. The exponential sensitivity of the cluster mass function to mass and redshift makes cluster counts and their evolution with redshift a very powerful probe of cosmological parameters (see Fig. 1.6). On the other hand, any systematic errors in the measurement of cluster mass are also exponentially amplified by the steepness of the mass function.

In order to measure the mass function using a large sample of clusters, we need to relate cluster mass to an easily observable quantity. One of the most diffuse techniques is to measure the temperature of the X-ray emitting gas, because it is closely related to the depth of a cluster's potential well and can be readily observed to $z \sim 1$ with current X-ray telescopes. Given the precision reached in the observations, systematic uncertainty in the link between mass and temperature is the main factor limiting this technique. Although mass and temperature are simply related for a cluster in hydrostatic equilibrium, this condition is often not met in real objects, where the turbulent velocities can sometimes be 20-30% of the sound speed. In this case, the hydrostatic assumption would lead to masses underestimated by 10-15% (Rasia et al., 2004).

Another possible proxy for the mass of a cluster is its galactic content. Optical telescopes have gathered much larger cluster samples than have X-ray telescopes, but deriving a mass function from these samples is not so

straightforward, because of projection effects and of the difficulty to measure the effective volume associated with a given cluster mass in a richness-selected survey. From the theoretical point of view, there is the need to calibrate the light-to-mass ratio, which introduces another uncertainty. For these reasons, optical richness alone is not a competitive method to weigh large samples of clusters for cosmological purposes.

Weak lensing is a very promising method for measuring cluster masses, because, as we will see later, it is independent of a cluster's baryon content, dynamical state, and mass-to-light ratio. The main systematic problem in weak-lensing mass measurements comes from lensing produced by mass outside the virial radius but along the line of sight through the cluster.

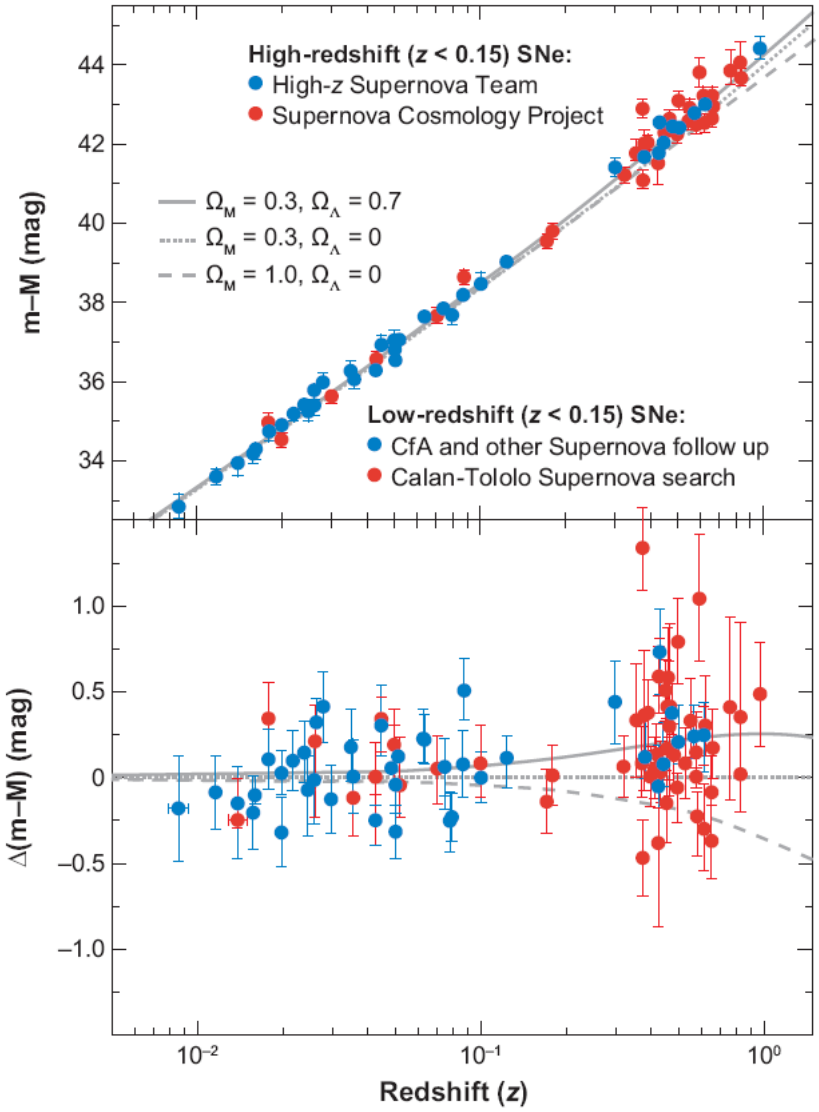

Figure 1.1: Distance modulus (m-M) as a function of redshift for Supernovae Ia. Figure taken from Frieman et al. (2008).

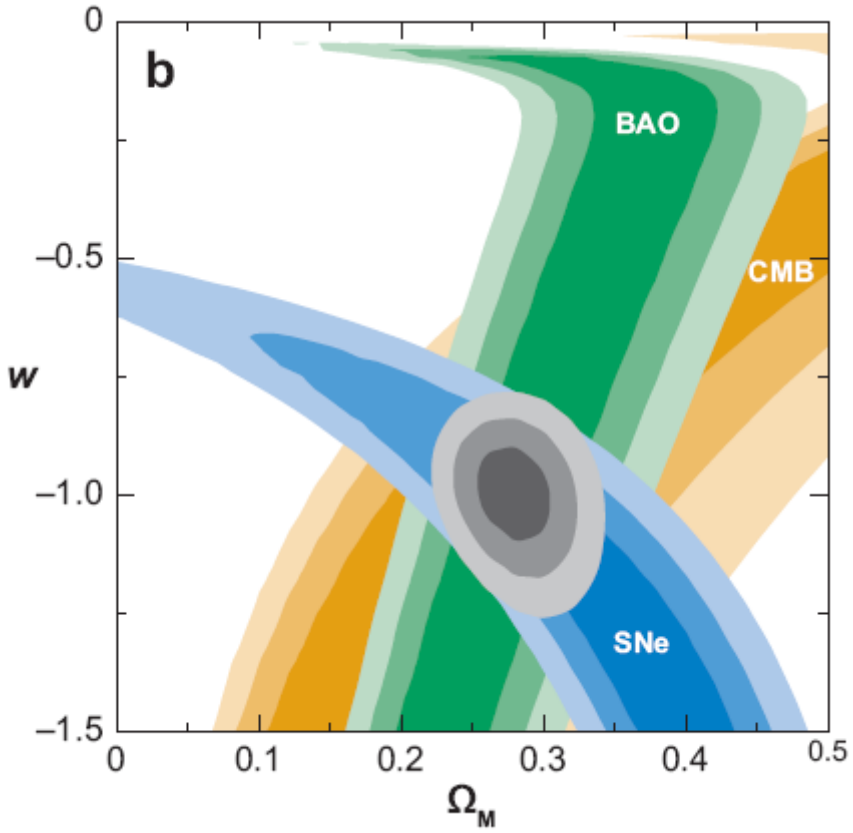

Figure 1.2: Constraints on Ω_M and the equation-of-state parameter w from different observables. Figure taken from Frieman et al. (2008).

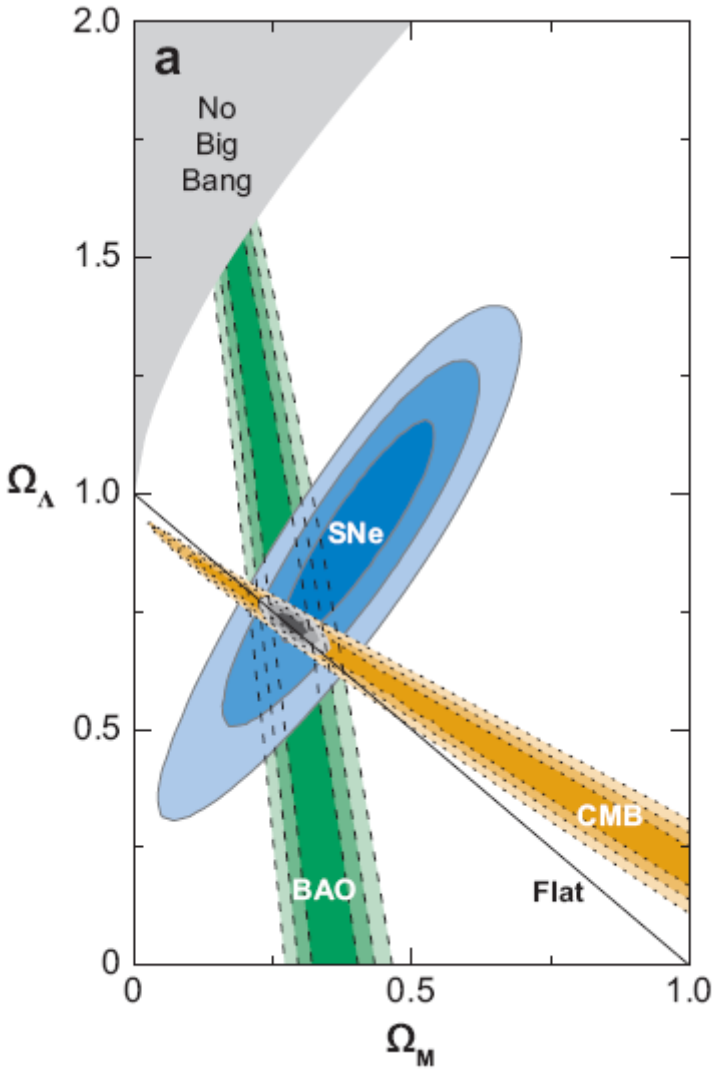

Figure 1.3: Constraints on Ω_M and Ω_Λ from different observables. Figure taken from Frieman et al. (2008).

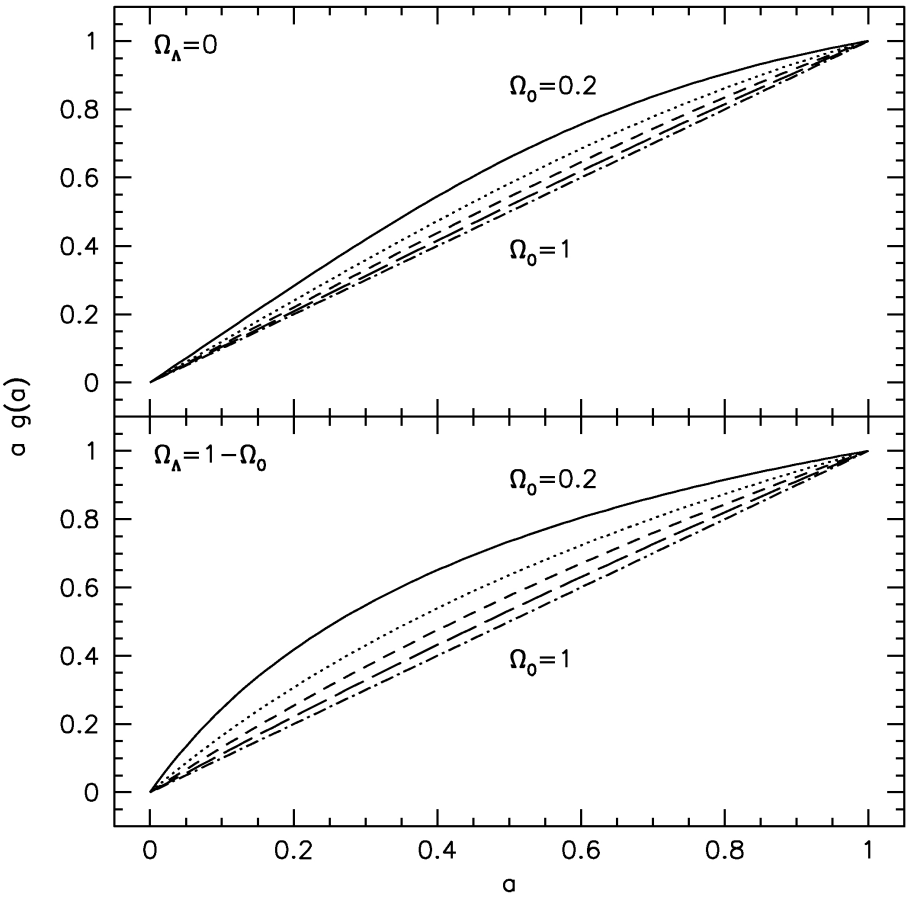

Figure 1.4: The function $ag(a)$ introduced in Eqs. (1.48) and (1.49) for $\Omega_0 = \Omega_{0,M}$ variable between 0.2 and 1. Taken from Bartelmann & Schneider (2001)

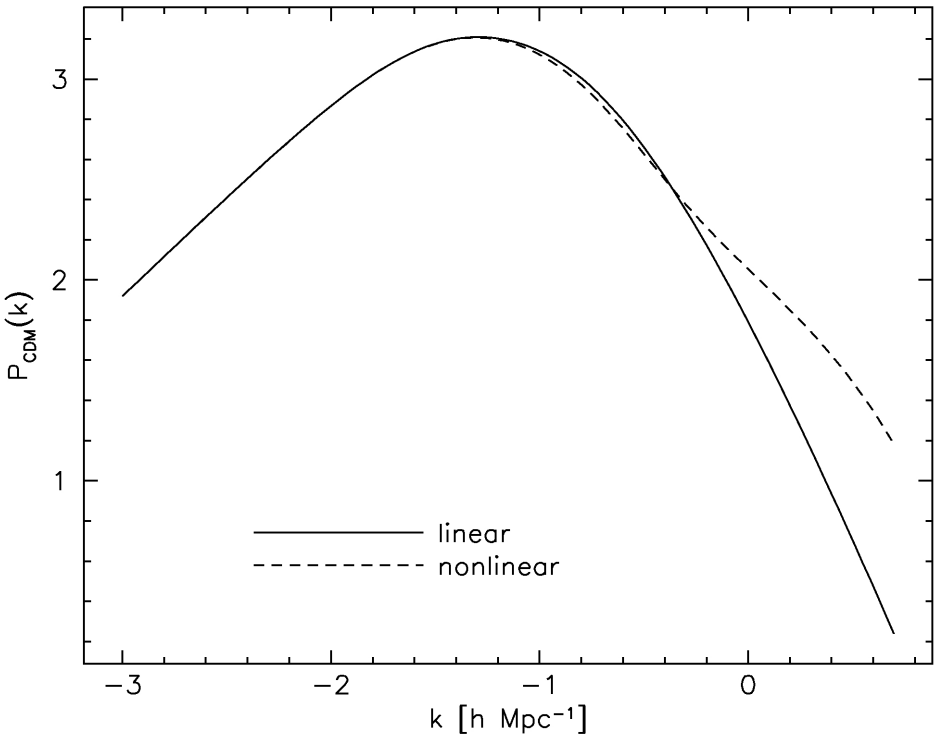

Figure 1.5: Theoretical CDM power spectrum at t_0, in a flat model without cosmological constant, normalised according to the local density of cluster of galaxies. The influence of the non-linear growth of perturbations can be seen at small scales, where the spectrum detaches from $\propto k^{-3}$. Taken from Bartelmann & Schneider (2001).

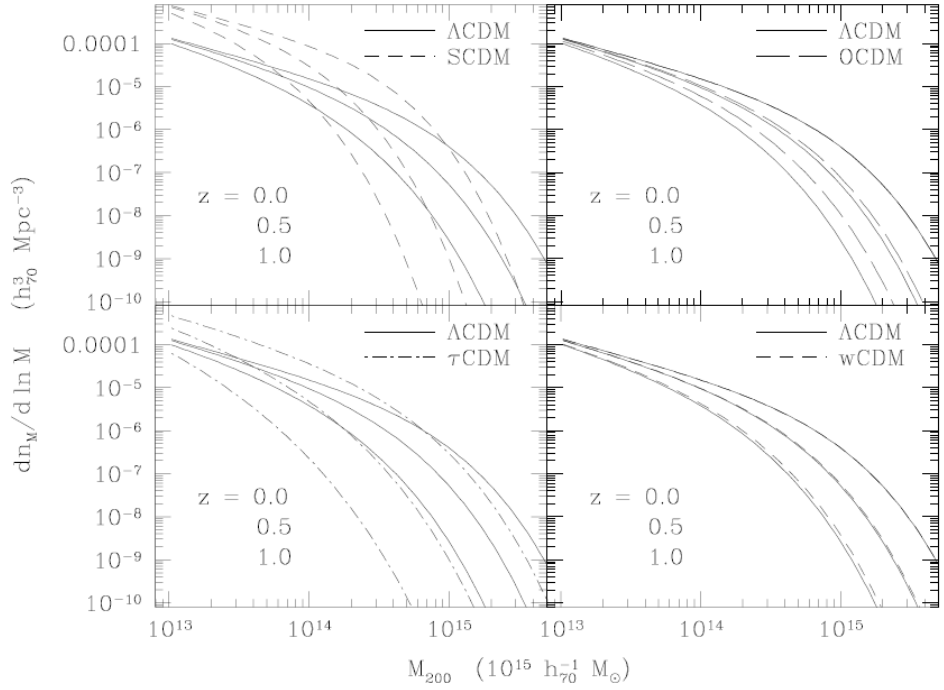

Figure 1.6: Mass-function evolution in five different cosmologies. The fiducial model in all cases is the ΛCDM model with $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, w = -1, and $\sigma_8 = 0.9$. The upper left panel compares cluster evolution in the ΛCDM case with a standard cold dark- matter model (SCDM) having $\Omega_M = 1.0$, $\Omega_\Lambda = 0.0$, and $\sigma_8 = 0.5$. Retaining $\Omega_M = 1.0$ and $\Omega_\Lambda = 0.0$ while adjusting the power spectrum so that $\Gamma = \Omega_M h \exp[-\Omega_b(1 + \sqrt{2h}/\Omega_M) = 0.21$ gives a τCDM model (lower left). Dispensing with dark energy while keeping the matter density low gives an OCDM model ($\Omega_M = 0.3$, $\Omega_\Lambda = 0$, $\sigma_8 = 0.9$; upper right). In the lower right panel, it is considered a model (ωCDM) identical to ΛCDM, except for ω = -0.8.

Chapter 2

Weak lensing surveys

2.1 Gravitational lensing

2.1.1 Deflection angle

Light travels in the space-time along null geodesics. In an ideal perfectly homogeneous Universe, photon would follow straight trajectories. The presence of inhomogeneities and structures perturbs the space-time and forces photons to deviate from their initial direction. This phenomenon is named *gravitational lensing*. The most simple cases is that of a single gravitational potential Φ embedded in a homogeneous Universe. If the potential is relatively small ($\Phi \ll c^2$) and moving much slower than the speed of light, it deflects the photon trajectory by an angle

$$\vec{\hat{\alpha}} = \frac{2}{c^2} \int \vec{\nabla}_\perp \Phi dl \,, \tag{2.1}$$

where the notation $\vec{\nabla}_\perp$ means that the gradient must be computed perpendicularly to the light path. In this simple case, one can perform the so-called *thin lens approximation*, that describes the trajectory of the photon as a broken line, that changes direction only once passing next to the location of the potential. The real angular position of the source and the one that appears to the observer are connected through the *lens equation*. For convenience, one defines a plane at the source distance, named *source plane*, and a parallel one where the deflection happens, named *lens plane*. We call D_D the angular diameter distance of the lens, D_S the one of the source, D_{DS} the

distance between the source and the lens. We remind that, for a general (not flat) space-time, $D_{DS} \neq D_S - D_D$. The two-dimensional vector $\vec{\eta}$ represents the position on the source plane, $\vec{\xi}$, the one on the lens plane; $\vec{\beta} = \vec{\eta}/D_S$ e $\vec{\theta} = \vec{\xi}/D_D$ are the corresponding angular coordinates. From Figure 2.1 one derives the geometrical relation

$$D_S\vec{\theta} = D_S\vec{\beta} + D_{DS}\vec{\hat{\alpha}}(D_D\vec{\theta})\,, \tag{2.2}$$

where $\vec{\hat{\alpha}}$ is again the deflection angle. Although $\vec{\hat{\alpha}}$ is the physical quantity

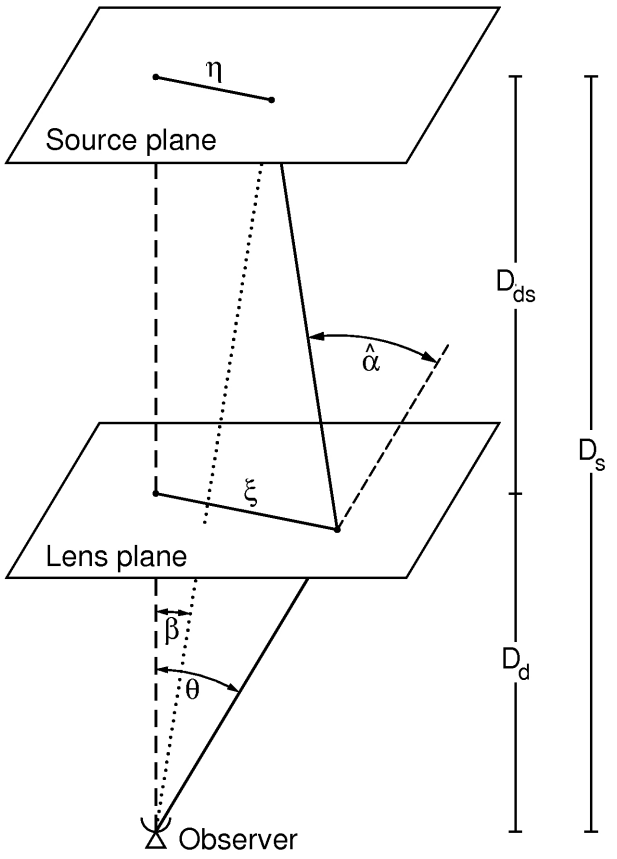

Figure 2.1: Schematic representation of the gravitational lensing in the thin lens approximation. Figure taken from Bartelmann & Schneider (2001)

affected by lensing, it is not an observable. So it is common to define the

reduced deflection angle $\vec{\alpha}$ as

$$\vec{\alpha}(\vec{\theta}) = \frac{D_{DS}}{D_S}\vec{\hat{\alpha}}(D_D\vec{\theta})\,, \tag{2.3}$$

that corresponds to the angular difference between the observed source position and the real one. In fact, from Eqs. 2.1 and 2.2, one obtains

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta})\,, \tag{2.4}$$

that is the lens equation we were looking for.

As noted already in Eq. 2.1, the lensing effect of a single lens depends on the gravitational potential integrated along the line of sight. This can be formalised by defining the *lensing potential* Ψ as

$$\Psi(\vec{\theta}) = \frac{D_{DS}}{D_D D_S}\frac{2}{c^2}\int \Phi(D_D\vec{\theta}, z)dz\,. \tag{2.5}$$

Now, the (reduced) deflection angle is equal to the gradient of Ψ

$$\nabla_\theta\Psi = \frac{D_{DS}}{D_S}\frac{2}{c^2}\int \nabla_\perp\Phi(\vec{\xi}, z)dz = \frac{D_{DS}}{D_S}\vec{\hat{\alpha}} = \vec{\alpha}\,. \tag{2.6}$$

2.1.2 Shapes of extended sources

As the deflection angle changes as a function of the position on the source plane $\vec{\eta}$, the shape of an extended source is not conserved by gravitational lensing. In principle, the lens equation should be solved for any infinitesimal element of the image to calculate the original shape of the source. In case the scale over which the properties of the lens change is much larger than the source image, a linearisation of the lens equation around the center of the image is enough to understand the effects of lensing. The jacobian matrix A_{ij} is defined as

$$A_{ij} \equiv \frac{\partial\beta_i}{\partial\theta_j}\,. \tag{2.7}$$

Remembering Eq. 2.4, it becomes

$$A_{ij} = \delta_{ij} - \frac{\partial\alpha_i}{\partial\theta_j} = \delta_{ij} - \frac{\partial^2\Psi}{\partial\theta_i\partial\theta_j}\,. \tag{2.8}$$

The matrix A can be written as

$$A = \begin{pmatrix} 1-\Psi_{11} & -\Psi_{12} \\ -\Psi_{12} & 1-\Psi_{22} \end{pmatrix}\,, \tag{2.9}$$

where Ψ_i stands for $\partial\Psi/\partial\theta_i$. Defining the *convergence* κ as

$$\kappa = \frac{1}{2}\nabla^2\Psi = \frac{1}{2}(\Psi_{11} + \Psi_{22}) \tag{2.10}$$

and introducing the two components of the *shear* $\gamma = \gamma_1 + i\gamma_2$,

$$\gamma_1 = \frac{1}{2}\left(\frac{\partial^2\Psi}{\partial\theta_1^2} - \frac{\partial^2\Psi}{\partial\theta_2^2}\right) = \frac{1}{2}(\Psi_{11} - \Psi_{22})\,, \tag{2.11}$$

$$\gamma_2 = \frac{\partial^2\Psi}{\partial\theta_1\theta_2} = \Psi_{12}\,, \tag{2.12}$$

Eq. 2.9 can be rewritten as

$$A = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{2.13}$$

The effect of the matrix A on the observed shape of the image can be clarified by writing it as the sum of two parts:

$$A = (1-\kappa)\mathcal{I} - \gamma \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}, \tag{2.14}$$

where $\mathcal{I}$ is the identity matrix and we introduced 2ϕ as the argument of the complex shear,

$$\gamma_1 = \gamma\cos 2\phi, \quad \gamma_2 = \gamma\sin 2\phi\,. \tag{2.15}$$

From Eq. 2.14 we see that the convergence produces an isotropic distortion of the source, whilst the shear contribution is anisotropic, as shown in Fig. 2.2. In particular one can show that a lens with $\gamma \neq 0$ transforms a circular source into an ellipse with axes proportional to $(1-\kappa\pm\gamma)^{-1}$ in the direction ϕ.

2.1.3 Lensing phenomenology

The astronomical realizations of gravitational lensing are usually split in two broad classes: *strong lensing* happens when an object is double imaged or strongly distorted by a lens, while the name *weak lensing* applies to small deviations in the observed shape of astronomical objects that can be detected only via statistical procedures.

Examples of strong lensing features are:

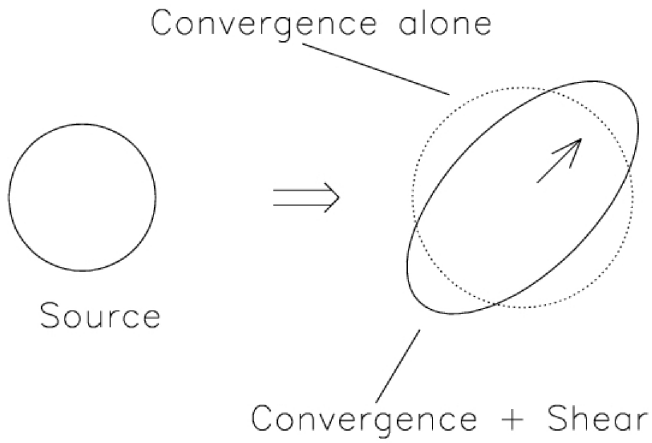

Figure 2.2: Effects of convergence and shear on an idealised galactic image. Figure taken from Narayan & Bartelmann (1996).

- multi-imaged quasars, that were the first extragalactic lensing phenomenon do be recognised. From the time delay between the luminosity variations of the different images, it is possible to draw estimates of H_0;

- Einstein rings, that happen when a source is exactly aligned with a spherically symmetric lens. The radius of the ring gives a good estimate of the enclosed mass of the lens;

- gravitational arcs, that form when the source is next to a *caustic*, the locus where the magnification is theoretically infinite. Also this elongated images can be used as tools to measure the matter distribution of the lens.

A special kind of strong lensing events are those in which the different images of the source cannot be resolved, because of the little angular separation, but nevertheless a variability in the source flux is detected. This happens for example when quasars are lensed by extragalactic stars, or when stars in the nearby galaxies are lensed by compact object in our galaxy. This phenomenon is named *microlensing*.

Weak lensing, instead, cannot be detected or studied through a single image, but changes the statistics of the galaxy population. The appearance of each galaxy is modified by the intervening matter, both in single massive objects like clusters of galaxies, or in the large-scale distribution of the Universe. This effect is usually negligible with respect to the intrinsic distribution in magnitude and ellipticity of the galaxy population. As we will see in the next Chapter, though, employing statistical methods it is possible to extract important information about the lens properties. Apart from shear, which we will focus on in the following, it is possible to detect lensing also through the *magnification bias*, that is the increased brightness of objects due to lens focusing. Anyway, this measure is much more difficult and it has been done in practice just a few times until now (see e.g. Hildebrandt et al., 2009).

2.1.4 Ellipticity and shear

If all galaxies were intrinsically circular, the ellipticity induced on them by lensing would directly tell us the second derivatives of the potential Φ. As instead galaxies have an intrinsic ellipticity distribution, meaningful information can be deduced only via statistical analysis over a large sample of them. In the following we will describe how to relate the observed shape of galaxies to the shear field. Let $I(\vec{\theta})$ be the surface brightness of a galaxy at the angular position $\vec{\theta}$. If the galaxy is isolated, we can measure I to large distances from the center $\bar{\vec{\theta}} = (\bar{\theta}_1, \bar{\theta}_2)$, defined as

$$\bar{\vec{\theta}} = \frac{\int d^2\theta q_I[I(\vec{\theta})]\vec{\theta}}{\int d^2\theta q_I[I(\vec{\theta})]}, \tag{2.16}$$

where $q_I[I(\vec{\theta})]$ is a weight function. For example, if $q_I = H(I - I_{th})$ is a Heaviside step function, $\bar{\vec{\theta}}$ is the center of the area delimited by the isophote I_{th}; else if $q_I = I$, then $\bar{\vec{\theta}}$ is the center of light. Once the proper $q_I[I(\vec{\theta})]$ has been chosen, the second brightness moment tensor is defined as

$$Q_{ij} = \frac{\int d^2\theta q_I[I(\vec{\theta})](\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int d^2\theta q_I[I(\vec{\theta})]}, \text{where } i, j \in 1, 2\,. \tag{2.17}$$

We can quantify its shape by the *complex ellipticity* ε

$$\varepsilon = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} - Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}}, \tag{2.18}$$

If the image has elliptical isophotes, with major axis a and minor axis b, the measured ellipticity is

$$\varepsilon = \frac{1 - b/a}{1 + b/a} e^{2i\vartheta} \tag{2.19}$$

where ϑ is the angular position of the major axis. According to this definition, the ellipticity is invariant under a rotation by an angle π. We can define the center of the source $\vec{\bar{\beta}}$ and its moments $Q_{ij}^{(s)}$ in analogy with Eqs. (2.16) and (2.17). From the jacobian matrix A (2.7), we get that

$$Q^{(s)} = AQA^T = AQA\,. \tag{2.20}$$

We can then define the complex ellipticity of the source $\varepsilon^{(s)}$ in analogy with Eq. 2.18. In this case the transformation is

$$\varepsilon^{(s)} = \begin{cases} \dfrac{\varepsilon - g}{1 - g^*\varepsilon} & \text{for } |g| \leq 1, \\ \\ \dfrac{1 - g\varepsilon^*}{\varepsilon^* - g^*} & \text{for } |g| > 1. \end{cases} \tag{2.21}$$

where we introduced the *reduced shear* g

$$g = \frac{\gamma}{1 - \kappa}\,. \tag{2.22}$$

The relations between ε and $\varepsilon^{(s)}$ show that the resulting shape depends on the reduced shear only. Thus, measures of ellipticity allow to estimate the complex shear, and not the shear and the convergence separately. This can be seen if we rewrite the jacobian matrix as

$$A = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}. \tag{2.23}$$

The $(1 - \kappa)$ factor changes the size of the image, but leaves untouched its shape. In the *weak lensing regime*, where $k \ll 1$ and $|\gamma| \ll 1$ (and thus $g \ll 1$), the first part of Eq. 2.21 reduces to $\varepsilon \simeq \varepsilon^{(s)} + g$.

If we assume to have a sample of sources at the same redshift, and that the expectation value of their source ellipticity vanishes,

$$\langle E(\varepsilon^{(s)}) \rangle = 0\,, \tag{2.24}$$

the expectation value of the measured ellipticity is

$$E(\varepsilon) = \begin{cases} g & \text{if } |g| \leq 1 \\ 1/g^* & \text{if } |g| > 1, \end{cases} \tag{2.25}$$

In the weak lensing regime, $|g| \ll 1$, and thus $\gamma \sim g \sim \langle\varepsilon\rangle$.

The previous result can be extended in case the sources are distributed in redshift. In that case, the expectation value depends on an integral over the redshift of the lens properties multiplied by a weighting function that depends on the redshift distribution of the sources.

The previous definitions cannot be applied directly when analysing real observations. For example, if we use the weighting function $q_I(I) = IH(I - I_{th})$, the choice I_{th} that delimits the object depends on the noise present in each pixel, that would influence the shear measurement in a very nonlinear way. It is common to substitute $q_I(I)$ with $IW(\vec{\theta})$, where $W(\vec{\theta})$ depends explicitly on $\vec{\theta}$. Another problem to face when dealing with real observations is the atmospheric turbulence, that translates a point-like source in an extended image. If $I(\vec{\theta})$ is the surface brightness of the source, the observed surface brightness is

$$I^{(obs)}(\vec{\theta}) = \int d^2\vartheta I(\vec{\vartheta}) P(\vec{\theta} - \vec{\vartheta}), \tag{2.26}$$

where $P(\vec{\theta})$ is the *point-spread function* (PSF) that describes the appearance on the CCD of a point-like source. The effect of the PSF can be separated in two parts: it makes small images more roundish and it introduces a systematical ellipticity in the measurements.

To extract shear information from real images, different methods can be employed. They can be roughly divided in two classes: passive methods, which measure the quadrupole moments of the galaxy image and then correct for the PSF (see e.g. Kaiser et al., 1995) and active methods, which fit the image with the result of the convolution between a galaxy model and the observed PSF (see e.g. Kitching et al., 2008). Their performances are usually summarised through

$$g^m - g^t = mg^t + c\,, \tag{2.27}$$

where g^t and g^m are the true and measured shear, respectively, while m and c are the multiplicative and additive errors. The performances of the methods

can be tested analysing simulations of observations with known input shear values, as is shown in Massey et al. (2007) and Bridle et al. (2010). Different methods are able to reach sub-percent values of the multiplicative bias m and an additive bias below 0.001.

2.2 Cosmic shear

2.2.1 Deflection by matter distribution

Apart from single massive objects, such as galaxies or clusters, the whole matter distribution produces distortions in the space-time, and thus deflects photon paths. The effect of this phenomenon on the observed shapes of galaxies is named *cosmic shear*.

To study cosmic shear, we need to introduce the general treatment of light propagation in curved space-times. One can show that in a homogeneous and isotropic Universe the comoving angular separation $\vec{x} = \vec{\xi}/a$ between two light rays varies according to

$$\frac{d^2\vec{x}}{d\omega^2} + K\vec{x} = 0\,. \tag{2.28}$$

The solution depends on the geometric parameter K. In the flat case, $K = 0$, the separation is a linear function of the comoving distance ω; in general, solutions are trigonometric (if $K > 0$) or hyperbolic (if $K < 0$) functions. We now add small ($\Phi \ll c^2$) and slow ($v \ll c$) perturbations on this idealised smooth Universe. If the length scales relevant for the perturbations are smaller than the curvature scale of the Universe, we can assume the background geometry as flat when the deviations due the inhomogeneities take place. In analogy with Eq. 2.1, the variation in the photon trajectory due to a potential Φ is

$$\frac{d^2\vec{x}}{d\omega^2} = -\frac{2}{c^2}\nabla_\perp\Phi\,. \tag{2.29}$$

On top of considering the general geometry of the space-time, another difference occurs with respect to the single-lens approximation. It is no longer interesting to consider the deflection angle with respect to an idealised unperturbed trajectory, as every photon will suffer many deflections along its path. Thus, we instead measure the variation of the angular distance between two

neighbouring rays as a function of the comoving distance from the observer. Adding up Eqs. (2.28) and (2.29), we obtain

$$\frac{d^2\vec{x}}{d\omega^2} + K\vec{x} = -\frac{2}{c^2}\Delta\{\nabla_\perp \Phi[\vec{x}(\vec{\theta},\omega)\,,\omega]\}\,, \tag{2.30}$$

where we introduced the difference Δ between the deflections suffered by the two rays. Solving Eq. 2.30, we get that the angular separation between the two rays as a function of distance is

$$\vec{x}(\vec{\theta},\omega) = f_K(\omega)\vec{\theta} - \frac{2}{c^2}\int_0^\omega d\omega' f_K(\omega-\omega')\Delta\{\nabla_\perp \Phi[\vec{x}(\vec{\theta},\omega),\omega]\}\,, \tag{2.31}$$

where the first term is due to the geometry of the smooth background space-time and follows the definition of the angular-diameter distance $f_K(\omega)$, while the second one is an integral over the different deflections taken by the light rays. Assuming that the variation of the angular separation between the two rays is small with respect to the angular separation of the unperturbed rays, the integral can be evaluated on the unperturbed trajectory $f_K(\omega')\vec{\theta}$, instead that on $\vec{x}(\vec{\theta},\omega')$. Now we can write the deflection angle $\vec{\alpha}$ at a distance ω, defined as

$$\vec{\alpha}(\vec{\theta},\omega) = \frac{f_K(\omega)\vec{\theta} - \vec{x}(\vec{\theta},\omega)}{f_K(\omega)}\,, \tag{2.32}$$

that is the separation between the two rays in a smooth universe minus their real separation, divided by the angular diameter distance. We get

$$\vec{\alpha}(\vec{\theta},\omega) = \frac{2}{c^2}\int_0^\omega d\omega' \frac{f_K(\omega-\omega')}{f_K(\omega)}\nabla_\perp \Phi[f_K(\omega')\vec{\theta},\omega']\,. \tag{2.33}$$

Because of its definition, this angle is not unique. All the physical quantities depend on derivatives of $\vec{\alpha}$.

2.2.2 Convergence power spectrum

Analogously to the single lens case, we can define an *effective convergence* κ_{eff} as

$$\kappa_{eff}(\vec{\theta},\omega) = \frac{1}{2}\nabla_\theta\cdot\vec{\alpha}(\vec{\theta},\omega) = \frac{1}{c^2}\int_0^\omega d\omega' \frac{f_K(\omega-\omega')f_K(\omega')}{f_K(\omega)}\frac{\partial^2}{\partial x_i \partial x_i}\Phi[f_K(\omega')\vec{\theta},\omega']\,. \tag{2.34}$$

The derivatives are on the directions perpendicular to the line of sight. In order to make use of the cosmological Poisson equation,

$$\nabla^2\Phi = \frac{3H_0^2\Omega_0}{2a}\delta\,, \tag{2.35}$$

we can add the derivative of the lensing potential along the line of sight, as it goes to zero in our approximation. Thus we obtain

$$\kappa_{eff}(\vec{\theta},\omega) = \frac{3H_0^2\Omega_0}{2c^2}\int_0^{\omega} d\omega' \frac{f_K(\omega-\omega')f_K(\omega')}{f_K(\omega)}\frac{\delta[f_K(\omega')\vec{\theta},\omega']}{a(\omega')}\,, \tag{2.36}$$

where the effective convergence is linked to the perturbations encountered along the photon path. As every measurement regarding weak lensing is obtained by averaging over the line of sight, it is useful to define $\bar{\kappa}_{eff}$ as the mean value of κ_{eff} with respect to ω, weighted according to a source distribution $G(\omega)$,

$$\bar{\kappa}_{eff}(\vec{\theta}) = \int_0^{\omega_H} d\omega G(\omega)\kappa_{eff}(\vec{\theta},\omega)\,, \tag{2.37}$$

where ω_H is the horizon distance. Inserting Eq. 2.36, one can rewrite

$$\bar{\kappa}_{eff}(\vec{\theta}) = \frac{3H_0^2\Omega_0}{2c^2}\int_0^{\omega_H} d\omega \bar{W}(\omega)f_K(\omega)\frac{\delta[f_K(\omega)\vec{\theta},\omega]}{a(\omega)}\,, \tag{2.38}$$

where $\bar{W}(\omega)$ is the function that weighs perturbations at distance ω,

$$\bar{W}(\omega) = \int_w^{w_H} d\omega' G(\omega')\frac{f_K(\omega'-\omega)}{f_K(\omega')}\,. \tag{2.39}$$

As the mean effective convergence $\bar{\kappa}_{eff}$ is a weighted mean over the z dimension of the density contrast δ, one can use the Limber's equation to link the 2-d spectrum of the former to the 3-d spectrum of the latter. The result is:

$$P_\kappa(l) = \frac{9H_0^4\Omega_0^2}{4c^2}\int_0^{w_H} d\omega \frac{\bar{W}^2(\omega)}{a^2(\omega)}P_\delta\left(\frac{l}{f_K(\omega)},\omega\right). \tag{2.40}$$

Cosmology role in Eq. 2.40 is twofold. The background evolution of the Universe influences the way angular modes l are related to linear scales at different redshifts, while the growth of structure defines the power spectrum P_δ.

Convergence is a very difficult quantity to measure, so lensing observations usually focus on shear measurements. Luckily, the shear and the convergence, being both second-order derivatives of a potential Ψ, share the same power spectrum. Thus, measuring shear estimates over a patch of the sky, one can extract information about cosmology through Eq. 2.40. In practice, different statistics can be applied to data. The most common one is the two-point correlation function,

$$\xi_\gamma(\phi) = \langle \gamma(\vec{\theta})\gamma^*(\vec{\theta}+\vec{\phi})\rangle \,. \tag{2.41}$$

Given that the correlation function is by definition the Fourier anti-transform of the power spectrum, we have

$$\xi_\gamma(\phi) = \int \frac{d^2l}{(2\pi)^2} P_\kappa(l) \exp(-i\vec{l}\cdot\vec{\phi}) = \int_0^\infty \frac{ldl}{2\pi} P_\kappa(l) J_0(l\phi) \,. \tag{2.42}$$

The integration over the angle enclosed by $\vec{l}$ and $\vec{\phi}$ gives rise to the Bessel function $J_0(x)$. If P_κ is made explicit through Eq. 2.40, one gets

$$\xi_\gamma(\phi) = \frac{9H_0^4\Omega_0^2}{4c^4} \int_0^{w_H} d\omega f_K^2(\omega) \frac{\bar{W}^2(\omega)}{a^2(\omega)} \times \int_0^\infty \frac{kdk}{2\pi} P_\delta(k,\omega) J_0[f_K(\omega)k\phi] \,. \tag{2.43}$$

The angular correlation function ξ_γ is thus an integral over the perturbation power spectrum, weighted by a window-function in the Fourier space.

Until now we have considered a single galaxy population with a angular distance distribution $G(\omega)$, or equivalently a redshift distribution $P(z) = dz/d\omega G(\omega)$. Sometimes we have information not only on the redshift distribution of the total population, but also on the redshifts of single galaxies. In this case, we can subdivide our sample into redshift bins and calculate the cross-correlation power spectrum $P_\kappa^{ij}(l)$, that is given theoretically by:

$$P_\kappa^{ij}(l) = \frac{9H_0^4\Omega_0^2}{4c^2} \int_0^{w_H} d\omega \frac{\bar{W}_i(\omega)\bar{W}_j(\omega)}{a^2(\omega)} P_\delta\left(\frac{l}{f_K(\omega)}, \omega\right) . \tag{2.44}$$

where the $\bar{W}_i$ function is derived from G_i as in Eq. 2.39. In this way, it is possible to measure the evolution of the cosmic structure, weighing the power spectrum as a function of redshift in different ways with different sets of objects, as can be seen in Fig. 2.3.

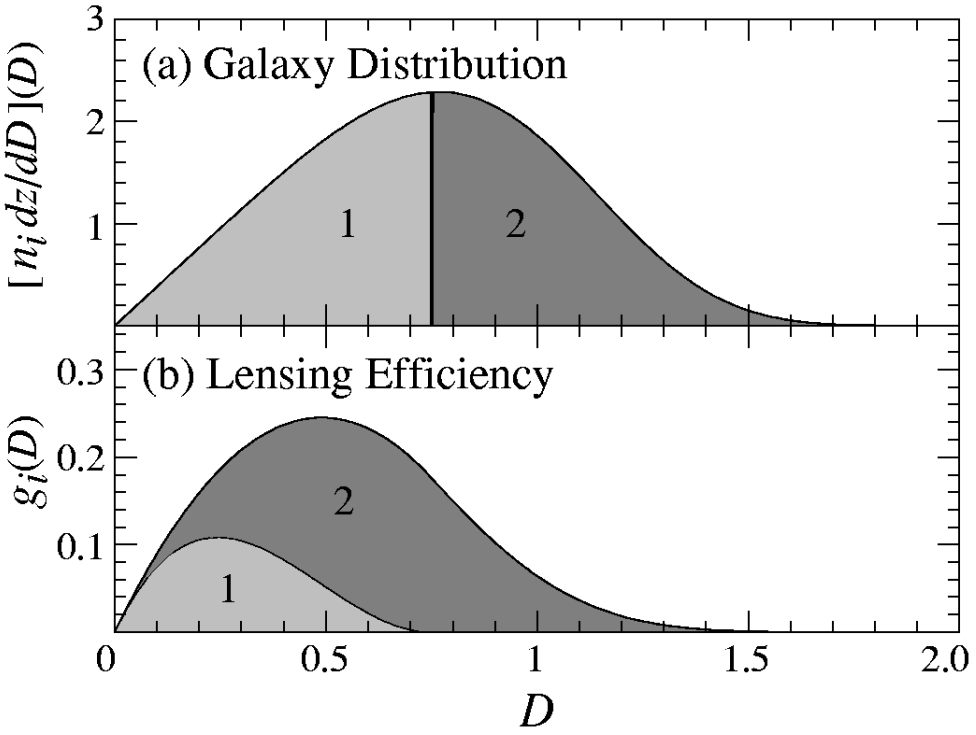

Figure 2.3: Sensitivity to the matter distribution as a function of redshift for two samples of galaxies. Figure taken from Hu (1999).

2.3 Observational uncertainties

2.3.1 Statistical noise

The distortion produced by the matter distribution on the galaxy observed shape is on average one order of magnitude smaller than the average galaxy ellipticity. This means that, on a single galaxy image, it is not possible to disentangle the lensing distortion from the intrinsic physical anisotropy of the object. The analysis must be made on a statistical basis, averaging over many observed objects. In practice, the considered observable is often the 2-point correlation function between the galaxy shear estimates, as described in Sect. 2.2. We assume that the correlation function is to be estimated in bins of angular width $\Delta\theta$. We define the function $\Delta_\theta(\phi)$, which is 1 for $\theta - \Delta\theta/2 < \phi \leq \theta + \Delta\theta/2$ and 0 otherwise. An estimator for the correlation

function $\hat{\xi}_+(\theta)$ is then

$$\hat{\xi}_+(\theta) = \frac{\Sigma_{ij}(\varepsilon_{it}\varepsilon_{jt} + (\varepsilon_{i\times}\varepsilon_{j\times})\Delta_{theta}\vec{\theta}(|\vec{\theta}_i - \vec{\theta}_j|)}{N_p(\theta)}\,, \tag{2.45}$$

where the number of pairs N_p is given by

$$N_p(\theta) = \Sigma_{ij}\Delta_\theta(|\vec{\theta}_i - \vec{\theta}_j|)\,, \tag{2.46}$$

and ε_t and $\varepsilon_\times$ are the tangential and cross components of the shear with respect to the line connecting the two galaxies. The expectation value of this estimator is obtained by averaging over the source ellipticities, assumed to be randomly oriented, and an ensemble average of the shear field. Since

$$\langle\varepsilon_{it}\varepsilon_{jt} + \varepsilon_{i\times}\varepsilon_{j\times}\rangle = \sigma_\varepsilon^2\delta_{ij} + \xi_+(|\vec{\theta}_i - \vec{\theta}_j|)\,, \tag{2.47}$$

where σ_ε is the dispersion of the intrinsic galaxy ellipticity, we see immediately that $\hat{\xi}_+(\theta)$ is an unbiased estimator of $\xi_+(\theta)$,

$$\langle\hat{\xi}_+(\theta)\rangle = \xi_+(\theta)\,. \tag{2.48}$$

The covariance $Cov(\hat{\xi}_+, \theta_1; \hat{\xi}_+, \theta_2)$ of this estimator is defined as

$$Cov(\hat{\xi}_+, \theta_1; \hat{\xi}_+, \theta_2) = \langle(\hat{\xi}_+(\theta_1) - \xi_+(\theta_2))(\hat{\xi}_+(\theta_2) - xi_+(\theta_2))\rangle\,. \tag{2.49}$$

Schneider et al. (2002) showed that its expectation value is

$$E(Cov(\hat{\xi}_+, \theta_1; \hat{\xi}_+, \theta_2)) = D\bar{\delta}(\theta_1 - \theta_2) + q_{++} + r_{+1} + r_{+0}\,, \tag{2.50}$$

where D is the only term that depends on the bin size and can be written as

$$D = 3.979{\times}10^{-9}\left(\frac{\sigma_\varepsilon}{0.3}\right)^4\left(\frac{A}{1deg^2}\right)^{-1}\left(\frac{n}{30arcmin^{-2}}\right)^{-2}\left(\frac{\theta}{1arcmin}\right)^{-2}\left(\frac{\Delta\theta/\theta}{0.1}\right)^{-1}, \tag{2.51}$$

where n is the galaxy number density and A is the survey area. The delta function $\delta(\theta_1 - \theta_2)$ means that this term matters only for the autovariance. The other three terms can be written as

$$q_{++} = \frac{2\sigma_\varepsilon^2}{\pi An}Q_{++} \tag{2.52}$$

$$r_{+1} = \frac{2}{\pi A}R_{+1} \tag{2.53}$$

$$r_{+0} = \frac{1}{2\pi A} R_{+0} \tag{2.54}$$

where Q_{++}, R_{+1} and R_{+0} are integrals of the correlation function. We note that only the autovariance is affected by the bin width $\Delta\theta$ and that all terms are proportional to the inverse of A. The last two terms, which are independent of σ_ε and n, represent the cosmic variance and set a limit for the precision on the evaluation of the shear correlation function independently of the quality of the observation.

2.3.2 Image quality

As we have shown in Sect. 2.2, to perform shear measurements it is fundamental to correct for the PSF. To perform such a correction, the PSF must be known with high precision at each point in the field-of-view of the instrument during the observing time. Paulin-Henriksson et al. (2008) have shown that the multiplicative bias induced by the PSF is

$$m = \frac{1}{P_\gamma}\left(\mu + \frac{\sigma[R^2_{PSF}]}{R^2}\right)\left\langle\left(\frac{R_{PSF}}{R_{gal}}\right)^2\right\rangle \tag{2.55}$$

for a galaxy population with radius R_{gal}. In the previous equation, $P_\gamma \sim 1.84$ is the shear polarisability and μ is a measure of the accuracy of the algorithm used to measure galaxy shapes. This term depends on the fractional uncertainty in the knowledge of the PSF size $\sigma(R^2_{PSF})/R^2_{PSF}$. Similarly for the same population of galaxies the additive bias is given by

$$\sigma^2_{sys} = \frac{1}{P_\gamma}\left\langle\left(\frac{R_{PSF}}{R_{gal}}\right)^2\right\rangle\left[\left(\frac{\sigma[R^2_{PSF}]}{R^2} + c\right)^2 |e_{PSF}|^2 + 2\sigma^2[e_PSF]\right] \tag{2.56}$$

which also depends on the absolute PSF ellipticity e_{PSF} and the uncertainty in the knowledge of the ellipticity $\sigma^2(e_{PSF})$. The parameter c is a measure of how well the algorithm used to measure galaxy shapes can correct for PSF anisotropy. These definitions enable us to relate the PSF properties directly to systematic biases: for example the larger R^2_{PSF} the larger the correction for the PSF blurring required.

2.3.3 Photometric redshifts

Photometric redshifts are crucial to perform weak lensing tomography. As to perform weak lensing analysis we need to observe a large amount of galaxies

to overwhelm the intrinsic ellipticity dispersion, it is virtually impossible to perform spectroscopic observations for the majority of them. Then, we must rely on photometric techniques that relate the observed flux of galaxies in a small (from 4 to ~ 30) number of bands to the redshift of the galaxy. To this end, different methods can be used and they can be subdivided in two groups: *template fitting* methods, which fit the observed flux of the galaxy with a set of spectra, and *training set* methods, which use a spectroscopic subsample to obtain empirical relations between fluxes and redshift. The former are more elastic, because they do not need a spectroscopic coverage of the survey, at the price of making harder assumptions on the galaxy population observed. Regardless of which technique is used, the efficacy of the method depends on the spectral type of the galaxy and the wavelength region covered by the observed wave bands. The primary redshift signal from photometric data arises from prominent breaks in the galaxy SED, such as the 400 nm Balmer break, and the Lyman break. Galaxies with large spectral breaks which are straddled by the filter bands (preferably with at least one band completely on each side of the break) will have the most robust and accurate measurements. For example, for data solely in the observed optical bands, early-type galaxies at low redshift ($z < 0.8$, beyond which the 400 nm break enters the I band), or star-bursting galaxies at high redshift ($z > 3$, where the Lyman break enters the U band) will have the strongest redshift signal. The typical redshift uncertainty is of the order of 0.01 - 0.1, depending on the depth and the band coverage of the survey. A catastrophic error in photometric redshift determination occurs when two templates with very different redshifts occupy regions in the multi-color space that are too close, so that for a variety of reasons, the template with the incorrect redshift is chosen. This could be the result of poor signal-to-noise ratio in the data, unrealistic templates being used, or simply a real degeneracy arising from the very coarse spectral information available within both the data and the templates. The simplest example of catastrophic error in a sample spanning a sufficiently large redshift range arises from the confusion of the Lyman break with the 400 nm break.

2.4 The Euclid mission

2.4.1 Characteristics

Euclid is an ESA-approved experiment that will take place in 2019-2022. It aims at mapping the geometry of the Dark Universe through two complementary probes: weak lensing and galaxy clustering. In order to do so, three observing instruments will be used: an imager in the visible band, a multi-band photometer and a spectrometer in the near infrared. In this work I will focus on the weak lensing observations (imaging and photometry).

Weak lensing observations from the ground are limited by the presence of the atmosphere, that blurs the galaxy images. Although we can take into account this effect in the analysis, the spatial and time variation of the atmosphere conditions, together with the finite number of stars that can be observed to map the PSF, make it impossible to go below $\sigma_{sys} \sim 10^{-4}$. With this precision, a survey that observes 15000 square degrees will be practically useless, because the systematic error will dominate the statistical error. Observing from space, on the other hand, permits to achieve a better control of the PSF, and a diffraction limited resolution. Moreover, from space it is possible to observe with no gaps the wavelength range between 1 and 2 microns, which is partially obscured from earth. This fact is fundamental for the galaxy clustering part of the experiment, as the H_α line lies in this wavelength range for galaxies with redshift between 0.7 and 2. The weak lensing experiment takes advantage of this feature with the multi-band near-infrared photometer, that will characterise the spectral-energy distribution of galaxies, and help in the determination of their photometric redshifts. The data of these bands, as we will see later, is very important especially for galaxies in the redshift range between redshifts 1 and 2.5.

2.4.2 Expected performances for cosmic shear

In the Wide Survey, that will cover 15000 deg^2 of extragalactic sky, the observing limit for a 10σ detection should be equivalent to a $mag_{RIZ} = 24.5$ for extended sources. This should give around 30 galaxies per $arcmin^2$ useful for weak lensing analysis, for an unprecedented total of 1.5 billion galaxies. In order to measure the shape of galaxies correctly, not only the PSF must

be stable and well-known, but it must also be significantly smaller than the galaxy size. With a PSF FWHM equal to 0.2", virtually no galaxies will be discarded because of this limit.

The three near-infrared bands should reach a 5σ depth for point sources with $mag_{YJH} = 24$. These data, together with their optical counterparts from ground-based collaborations, should be sufficient to fulfil the requirements about photometric redshift precision. To reach the desired accuracy on cosmology, the precision in the redshift measurement for each single galaxy must be such that $\sigma_z/(1+z) < 0.05$ and the mean redshift for each tomographic bin should be smaller then $0.002(1+z)$. In the following we will study in deep detail this aspect. The redshift distribution of the sources is expected to have a median z equal to 0.9.

To extract the cosmological signal through cosmic shear tomography (see Sect. 2.2.2), the galaxy sample will be divided into 10 redshift bins that will contain approximately the same number of galaxies. Considering the uncertainties described in the previous Section, one can derive the expected precision on cosmological parameter determination for the Euclid survey through a Fisher matrix analysis, as done in Amara & Refregier (2007). The derived precision on the dark energy equation of state determination is $\Delta_{w_p} = 0.015$ and $\Delta_{w_a} = 0.15$.

2.4.3 Secondary probe: galaxy clusters

The Euclid mission is driven by two observables, that are cosmic shear and galaxy clustering. The setup of the instrument and of the survey is aimed to give the best results for these two kinds of observations. Nonetheless, the Euclid high-quality data will be useful for many other astrophysical and cosmological subjects, such as galaxy evolution, type Ia supernovae, high-redshift Universe through galaxies and quasars, and so on. Here we will focus on galaxy clusters, whose abundance as a function of mass and redshift will also give additional cosmological constraints to the ones of the main science drivers of the mission.

Galaxy clusters can be detected in Euclid data in many ways, including filtering of the weak lensing signal or searching for close galaxies in the spectroscopic data. The main channel for cluster detection will nevertheless be the analysis of photometric data, as already performed in other surveys, such

as SDSS. Many different techniques can be used to this goal. In the following, we will come back to this subject and present an optimal filtering method. It is important to underline that the calibration of the mass-richness relation, one of the main uncertainties in the cosmological interpretation of optical cluster samples, will be helped by the presence of weak lensing data, that allow for an independent measure of the cluster mass.

The abundance of galaxy clusters brings information mainly about the distribution of the primordial fluctuations and about the physics of gravitational collapse. Their effect on the determination of the cosmological parameters can be seen in Fig. 2.4.

	Modified Gravity	Dark Matter	Initial Conditions	Dark Energy		
Parameter	γ	m_ν/eV	f_{NL}	w_p	w_a	**FoM**
Euclid Primary	0.010	0.027	5.5	0.015	0.150	430
Euclid All	0.009	0.020	2.0	0.013	0.048	1540
Euclid+Planck	0.007	0.019	2.0	0.007	0.035	4020
Current	0.200	0.580	100	0.100	1.500	~10
Improvement Factor	**30**	**30**	**50**	**>10**	**>50**	**>300**

Figure 2.4: Forecast on the precision in the cosmological parameters determination by Euclid. Figure taken from Laureijs et al. (2011)

Chapter 3

Simulation of observations

3.1 Introduction

The systematics described in the previous Chapter need to be properly evaluated to estimate the possible achievements of future weak lensing experiments. The measurement of the ellipticity and of the photometric redshift of the galaxies depend in a nonlinear way on the instrumental and survey characteristics, such as pixel size, PSF width and wavelength coverage. Thus, it is important to assess the systematics present in the procedure used to extract information from data by working on simulated images that mimic the ones obtained by observations. In the following Section, we will describe the code used in our work, which was first presented in Meneghetti et al. (2008).

3.2 Our code

3.2.1 Galaxy catalogue

The first step in performing a realistic simulation of observations is the creation of a source catalogue, that in our case will be composed of galaxies. For the topic we are covering, it is fundamental to have good morphological and photometric data for the galaxies, and a realistic distribution in luminosity, redshift and spectral type. To obtain this result one can use data taken from real observations, or extract them from some model distribution

for each quantity. Starting from a catalogue drawn from observations has the advantages of being realistic by construction and of retaining all correlations between different measures which are present in the real data. The price to pay is the limited number of available sources and the impossibility of extrapolating objects which lie beyond the capabilities of the survey, e.g. dimmer than the photometric limit.

For the work described in Sect. 5.1, we use the galaxy catalogue extracted from the Hubble Ultra-Deep Field (Coe et al., 2006). It is composed by more than 9000 objects with a spectral type and a photometric redshift assigned from 6-band photometry in the optical and near-infrared. To perform our simulations, we use the luminosity in the i band and the colour and k-correction derived from the appropriate spectral-energy distribution (SED). The SEDs assigned by Coe et al. (2006) to each object of the sample are taken from a catalogue with 8 members: 4 from Coleman et al. (1980) that represent ellipticals, spirals and irregulars, 2 from Kinney et al. (1996), for starburst galaxies, and 2 simple-stellar population models from Bruzual & Charlot (2003) to account for the bluest objects in the sample. Two interpolations between neighbouring SEDs are allowed. The complete set of SEDs is shown in Fig. 3.1.

Each galaxy in the sample has a morphological description via decomposition in a set of basis functions called *shapelets*, that were first introduced in Refregier (2003). They allow to store the characteristics of the appearance of the galaxy in a few coefficients, and to perform in a simple way modifications such as rotations, rescaling, shearing etc. The light distribution of each galaxy is written in terms of the basis functions $B_n(\vec{x}, \beta)$,

$$I(\mathbf{x}) = \sum_{n_1,n_2=0}^{\infty} I_{\mathbf{n}} B_{\mathbf{n}}(\vec{x} - \vec{x_c}; \beta)\,, \tag{3.1}$$

where $I_{\mathbf{n}}$ are the coefficients of the decomposition and β is a scale parameter. The functions $B_n(\mathbf{x}; \beta)$ are

$$B_{\mathbf{n}}(\vec{x}; \beta) = \beta^{-1}\Phi_{n_1}(\beta^{-1}x_1)\Phi_{n_2}(\beta^{-1}x_2), \tag{3.2}$$

where the one-dimensional functions $\Phi_n(x)$ are related to Gauss-Hermite polynomials $H_n(x)$,

$$\Phi_n(x) = [2^n \pi^{1/2} n!]^{-1/2} H_n(x) \exp(-x^2/2). \tag{3.3}$$

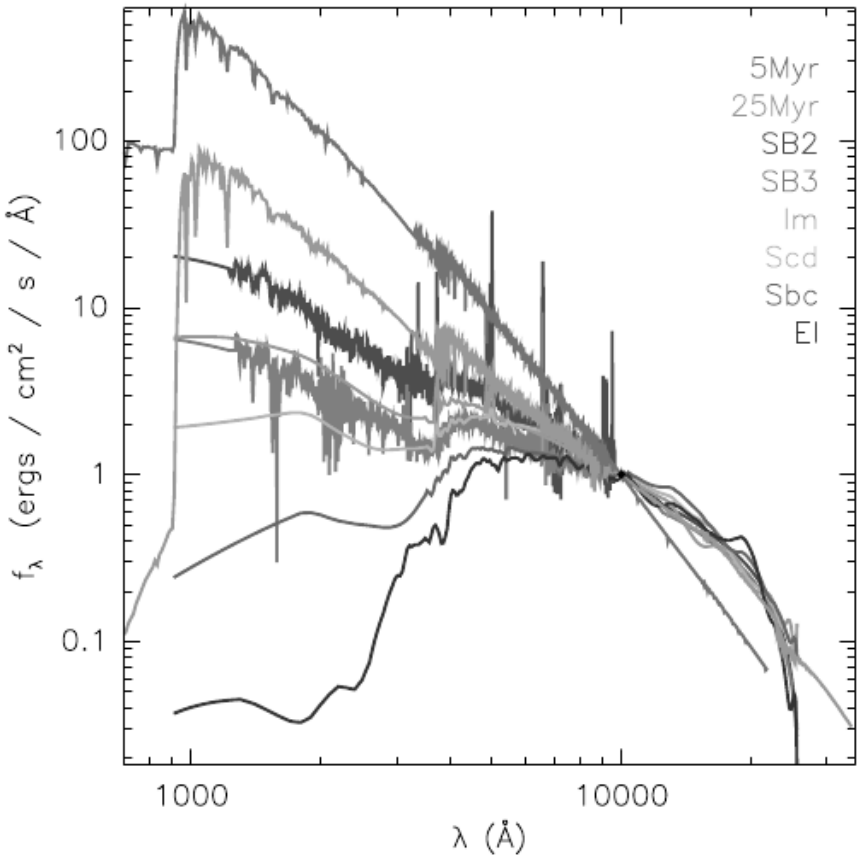

Figure 3.1: Set of template SEDs used for the galaxies in the HUDF catalogue. Figure taken from Coe et al. (2006).

The set of functions defined by Eq. 3.2, once fixed the scale parameter β, is a complete set of orthonormal functions, so that each continuous and integrable two-dimensional function $f(\mathbf{x})$, has one decomposition with coefficients $I_\mathbf{n}$,

$$I_\mathbf{n} = \int d^2x f(\mathbf{x}) B_\mathbf{n}(\mathbf{x};\beta)\,. \tag{3.4}$$

The shapelets of the three lowest orders are shown in Fig. 3.2.

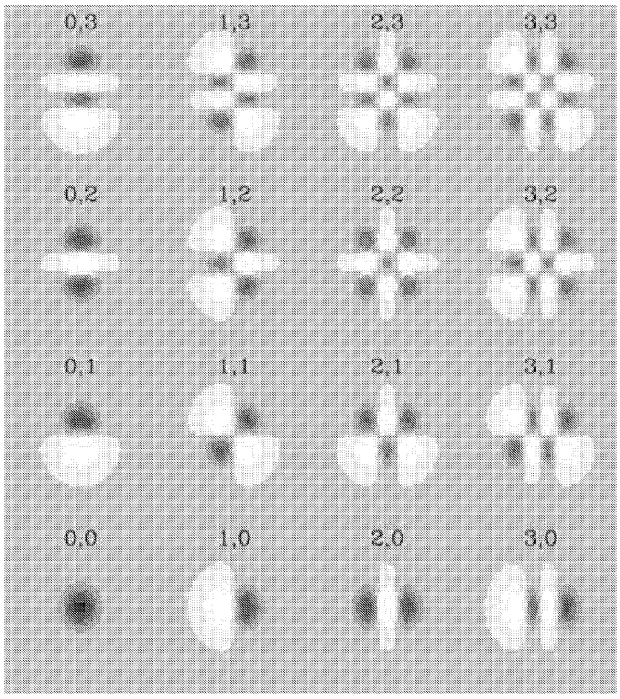

Figure 3.2: The 2-dimensional shapelets $B_n(\mathbf{x};\beta)$ of the three lowest orders, for fixed β. The light and dark regions correspond respectively to negative and positive regions. Figure taken from Refregier (2003).

3.2.2 Survey implementation

The implementation of an observation starts with the definition of the instrument that is used. The quantities that matter in the simulation are:

- the effective diameter of the mirror,
- the pixel size,
- the gain,
- the instrumental sources of noise, as read-out-noise and finite flatness of the field,
- the throughput curves of the mirror, the optics and the CCD.

The survey is defined by the exposure time, the field of view and the observing band. The quality of the image is also affected by the instrumental PSF, the atmospheric seeing and the background sky magnitude.

3.2.3 Image creation

We first determine the catalogue of sources we observe. If we start from an observed catalogue, we calculate the needed amount of galaxies by multiplying the field of view of our mock observation by the galaxy angular density of the original catalogue. Galaxies, in this case, are placed randomly in the observing field. For each galaxy, the luminosity in the given band is calculated considering its SED and the filter response as a function of wavelength. The shape of the galaxy is obtained by convolving its stored properties with the instrumental PSF. This operation is performed in the shapelets space. The relation between the coefficients of the convolved image $h_{\mathbf{n}}$ and the original ones $f_{\mathbf{n}}$ is

$$h_{\mathbf{n}} = \sum_{\mathbf{m},\mathbf{l}} C_{\mathbf{n},\mathbf{m},\mathbf{l}} f_{\mathbf{m}} g_{\mathbf{l}} \,, \tag{3.5}$$

where $g_{\mathbf{l}}$ are the coefficients of the function with which the convolution has to be performed. If we call α, β, e γ respectively the scale parameters of f, g e h, the C matrix can be written as the product of two one-dimensional convolution matrices

$$C_{\mathbf{n},\mathbf{m},\mathbf{l}}(\gamma,\alpha,\beta) = C_{n_1,m_1,l_1}(\gamma,\alpha,\beta) C_{n_2,m_2,l_2}(\gamma,\alpha,\beta) \,, \tag{3.6}$$

which are calculated as

$$C_{n,m,l}(\gamma,\alpha,\beta) = (2\pi)^{1/2}(-1)^n i^{n+m+l} \int_{-\infty}^{\infty} dx B_l(x,\gamma^{-1}) B_m(x,\alpha^{-1}) B_n(x,\beta^{-1}) \,. \tag{3.7}$$

A good choice for the scale parameter γ is given by $\gamma^2 = \alpha^2 + \beta^2$. Now the galaxy with the derived brightness distribution $I(\vec{x})$ can be observed. If the telescope has pixel size p and diameter D, the number of photons collected by the pixel $\vec{x}$ in the time t_{exp}, is given by (see Grazian et al., 2004)

$$n_\gamma(\vec{x}) = \frac{\pi D^2 t_{exp} p^2}{4h} I(\vec{x}) \int \frac{T(\lambda)}{\lambda} d\lambda \,, \tag{3.8}$$

where h is the Planck constant, $T(\lambda)$ is the total transmission, given by the product of the atmospheric extinction $A'(\lambda)$ and the efficiencies of CCD

$C(\lambda)$, filter $F(\lambda)$, mirror $M(\lambda)$ and optics $O(\lambda)$. The number of photons n_λ is then converted in ADU depending on the instrumental gain g

$$ADU(\vec{x}) = \frac{n_\gamma(\vec{x})}{g}\,. \tag{3.9}$$

To this ideal image of the galaxy, one must add the photons due to background light, calculated as

$$ADU_{sky} = \frac{\pi D^2 t_{exp} p^2}{4hg} \int \frac{T(\lambda)S(\lambda)}{\lambda} d\lambda\,, \tag{3.10}$$

where $S(\lambda)$ is the sky surface brightness. We assume that the number of photons collected by a pixel follows a Poissonian distribution, with expected value given by the sum of (3.9) and (3.10). The statistical and instrumental uncertainty is simulated adding a random quantity with Gaussian probability and standard deviation σ_N

$$\sigma_N(\vec{x}) = \left\{ n_{exp} \left(\frac{RON}{g} \right)^2 + (ADU(\vec{x}) + ADU_{sky}) + \left(f + \frac{A^2}{n_{exp}^2} \right) [ADU(\vec{x}) + ADU_{sky}]^2 \right\}^{1/2}, \tag{3.11}$$

where RON is the chip read-out-noise, n_{exp} is the number of exposures, A is the flat-field term. The quantity f measures the flat-field accuracy, that depends on the number of exposures and on the sky brightness as

$$f = (N_{ff} \cdot ADU_{sky} \cdot g)^{-1}\,. \tag{3.12}$$

Lensing effects are included by applying the lensing equation (2.4). Each pixel position $\vec{\theta}$ is connected to the position $\vec{\beta}$ on the plane where the ADU counts are assigned,

$$ADU_{lensed}(\vec{\theta}) = ADU[\beta(\vec{\theta})]\,. \tag{3.13}$$

In case of strong lensing, different coordinates $\vec{\theta}$ can refer to the same point $\vec{\beta}$ on the source plane.

3.3 Application to cluster mass estimation

As yet noted in Sect. 1.2.5, one of the biggest problems in using galaxy clusters as cosmological probes is the evaluation of their mass from observations. Both X-ray and lensing measurements suffer from systematics due

to the strong assumptions made to perform the calculations: hydrostatic equilibrium and sphericity. To evaluate properly the precision of the mass measurements made with these techniques, we perform simulations of observations both in the optical and in X-ray of clusters extracted from dynamical simulations. We summarise here this work, which is presented in Meneghetti et al. (2010).

Three clusters are observed along three orthogonal line of sights, for a total of nine objects to be analysed. For what concerns optical mock observations, we use the SkyLens code. For each cluster line of sight, a deflection map is created with a ray-tracing technique. Then, a proper amount of background galaxies are created according to the description of the previous Section. Two telescopes are simulated: HST and Subaru. For the SUBARU simulations, we assume an exposure time of 6000 s in the I band, with a seeing of 0.6". The PSF is assumed to be isotropic and modelled using a 2D Gaussian. For the HST simulations, we assume an exposure time of 7500 s with the F775W filter. An example of the simulated images can be seen in Fig. 3.3

The weak lensing measurements are done using the standard KSB method, proposed by Kaiser et al. (1995) and subsequently extended by Luppino & Kaiser (1997) and by Hoekstra et al. (1998). By selecting the galaxies with S/N > 10, we end up with catalogues of galaxy ellipticities containing ~ 30 sources/sq. arcmin. Different methods are used to extract a mass estimate from these catalogues: a fit of a NFW profile (Navarro et al., 1997), the aperture mass densitometry (Fahlman et al., 1994; Clowe et al., 1998), and a nonparametric reconstruction of the lens, where also strong lensing features are taken into account.

The comparison between the estimated mass from lensing and the true mass of the cluster are shown in Figs. 3.4 and 3.5. In the first case, the considered cluster mass is named 2D, because it is measured as all the matter was projected on the cluster lens plane. The accuracy of the mass estimates depends on the morphology of the lenses and on their substructures. Methods that assume that the lensing effect is due to a single mass clump tend to underestimate the mass when significant substructures are present. The nonparametric SL+WL method, on the other hand, does not require any assumption about the symmetry of the lensing signal, because substructures

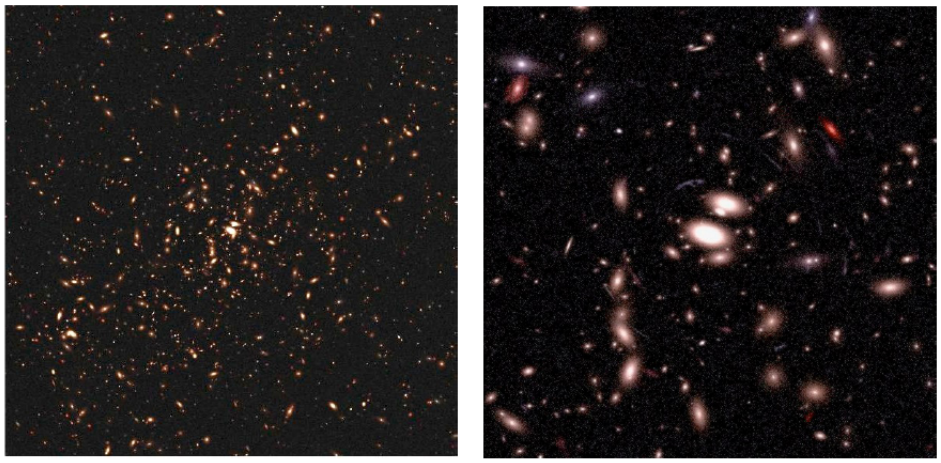

Figure 3.3: Left panel: color-composite image of a simulated galaxy cluster, obtained by combining three SUBARU exposures of 2500 s each in the B,V, I bands. The field-of-view corresponds to $\sim$ 450" $\times$ 450". Right panel: an HST/ACS composite image of the the central 100" $\times$ 100" of the same cluster. The image has been produced by combining mock observations with the filters F475W, F555W and F775W. Figure taken from Meneghetti et al. (2010).

are included in the mass model by construction. For this reason, the SL+WL method can recover the input mass with good precision even in the case of morphologically disturbed clusters. We find that the deviations between estimated and true masses are typically below the 10% level.

Lensing is sensitive by construction on the total mass projected on the lens plane. To extract a 3D mass from lensing measurements, it is needed to fit the derived mass distribution with a suitable function (a NFW profile in our case) and to calculate the mass from this distribution. The geometry between the main axis of the cluster and the line of sight becomes very important in this procedure: we find that, in the cases of good alignment (i.e. small angles) between the major axis of the cluster and the projection axis, the lensing masses tend to be systematically larger than the true masses, while the opposite occurs in those cases where the major axis is nearly perpendicular to the line of sight. The scatter in the mass measurement due to triaxiality is around 20%.

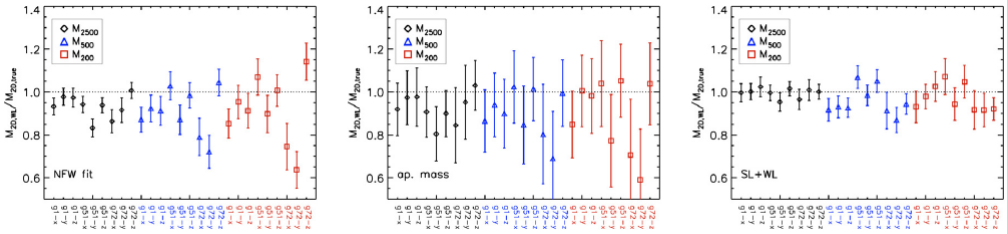

Figure 3.4: Comparison between the weak-lensing and the true 2D-masses of all the simulated clusters. From left to right, the panels refer to the methods based on the NFW fit of the shear profile, on the aperture mass densitometry, and on the non-parametric SL+WL reconstruction of the lensing potential, respectively. Shown are the ratios between the estimated and the true masses measured at three characteristic radii, namely R_{2500} (diamonds), R_{500} (triangles), and R_{200} (squares), versus the cluster names.

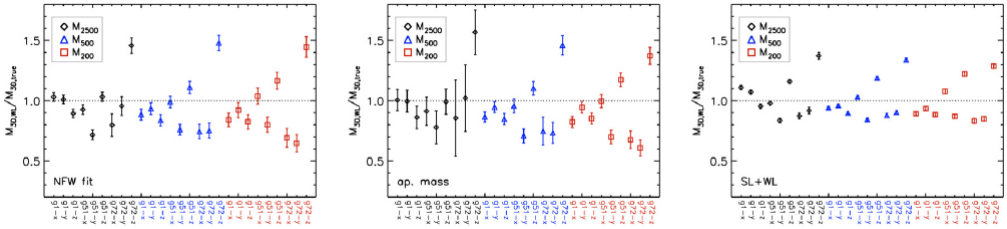

Figure 3.5: As in Fig. 3.4, but comparing the estimated and true 3D-masses.

X-ray measurements, in comparison, suffer less scatter but are usually biased low by ~ 10 %, because of pressure support induced by bulk motions. In general, the two measures can differ by up to 100 %, because they probe different physical quantities: the mass distribution projected along the line of sight in the case of lensing, and the depth of the gravity potential felt by the gas in case of X-ray observations. Interestingly, there are hints that the ratio of X-ray-to-lensing masses can be used as a tracer of hydrostatic equilibrium in the observed clusters.

Chapter 4

Cluster detection

4.1 Introduction

Clusters of galaxies lie in the densest regions of the Universe. In the standard cosmological model, they form from the highest peaks of the primordial density fluctuations. Their abundance as a function of mass and redshift is an extremely important tool to determine cosmological parameters and the history of structure formation. It is therefore fundamental to build robust tools to detect galaxy clusters in different observational bands.

In optical surveys, clusters of galaxies can be succesfully identified via overdensities of galaxies with respect to the mean field and via the coherent distortion of background galaxies due to gravitational lensing. These two observables come together with the same data but have very different properties. On one hand lensing analyses have the advantage of directly mapping dark matter structures, but to be performed they need high-quality imaging data, and their results are intrinsically very noisy. On the other hand, detecting clusters through galaxy overdensities is much less demanding from an observational point of view, because it does not require any shape measurement. Moreover, if an optical cluster identification is performed on a lensing survey field, it can be used as a useful tool to eliminate spurious detections.

A common approach in lensing is to convolve galaxy ellipticities with a suitable filter to find the cluster signal beyond the noise induced by intrinsic ellipticity of galaxies and cosmic shear by large-scale structures. Different filters have been proposed so far, including the Gaussian filter (Miyazaki et

al., 2002), the aperture mass in the version of Schneider et al. (1998), its modifications by Schirmer et al. (2004), and the one proposed by Hennawi & Spergel (2005). Maturi et al. (2005) proposed an optimal filter, which minimizes noises due to both the intrinsic ellipticity and the cosmic shear, selecting the scales on which the cluster signal-to-noise ratio is expected to be maximum. A comparison of the performance of the different filters used to find dark matter halos in cosmological simulations has been presented in Pace et al. (2007). An application to GaBoDs data of the optimal filter is in Maturi et al. (2007).

To build an optimal linear filter for galaxy overdensities, we start from the pioneering approach proposed by Postman et al. (1996), who developed an optical filter to analyse the data of the Palomar Distant Cluster Survey. Many authors have proposed different modifications of the original Postman filter: for example Kepner et al. (1999) introduced Poisson (instead of the Gaussian) statistics and a rough usage of photometric redshifts; Dong et al. (2008) allowed the radius of the filter to change to adapt it to the detected signal; Milkeraitis et al. (2009) tuned the parameters using mock light-cones extracted from the Millennium Simulation (Springel et al., 2005).

In the following, we construct a Postman-like filter as a linear filter, using the estimated richness of the cluster and not the likelihood of the data as main object of analysis. This allows us to keep the linearity with respect to the galaxy density and to calculate an error estimate analytically. We also introduce the information from photometric redshifts in the procedure.

The reference article for this Chapter is Bellagamba et al. (2011).

4.2 Optimal filtering: weak lensing

The linear matched filter for weak lensing cluster detections has been presented in Maturi et al. (2005). We refer the reader to that paper and to Pace et al. (2007) for a detailed description of its specific derivation, the comparison with other filtering techniques and tests of its performances. Here we summarise the main properties of the filter and how it can be applied to real data.

The observable quantity in weak lensing analyses is the reduced shear g (see Eq. 2.22), a measure of the distortion of background galaxies by inter-

vening structures (i.e. galaxy clusters). The weak lensing filter for cluster detection is expressed in Fourier space for convenience. Its shape is given by

$$\hat{\Psi}(k) = \frac{1}{(2\pi)^2}\left(\int \frac{\hat{\tau}^2(k)}{P_N(k)} d^2k\right)^{-1} \frac{\hat{\tau}(k)}{P_N(k)} \,, \tag{4.1}$$

where $\hat{\tau}$ is the Fourier transform of the weak lensing signal expected from a lensing cluster, i.e. the reduced shear, and P_N represents the power spectrum of the noise, due to the intrinsic ellipticity of the sources, their finite number on the sky and the additional shear induced by large-scale structures. The weak lensing signal model for clusters is computed assuming a spherically symmetric NFW density profile for the dark matter distribution. We also assume that the projected radial matter distribution of the cluster $\Sigma(x)$ follows an NFW model (Navarro et al., 1997; Bartelmann, 1996; Meneghetti et al., 2002):

$$\Sigma(x) = \frac{2\rho_s r_s}{x^2 - 1} f(x), \tag{4.2}$$

with

$$f(x) = \begin{cases} 1 - \frac{2}{\sqrt{x^2-1}} \arctan\sqrt{\frac{x-1}{x+1}} & (x > 1) \\ 1 - \frac{2}{\sqrt{1-x^2}} \operatorname{arctanh}\sqrt{\frac{1-x}{1+x}} & (x < 1) \\ 0 & (x = 1) \end{cases} ,$$

where $x \equiv r/r_s$. This distribution, apart from the normalization ρ_s, depends on the scale radius r_s only, corresponding to R_{200}/c, where R_{200} is the scale where the galaxy density is 200 times the critical density, and c is the so-called concentration parameter. The shape of the cluster signal, of the different sources of noise and of the resulting filter is shown in Fig. 4.1.

The estimate of the lensing signal is obtained by convolving the shear data, $D(\theta)$, with the filter Ψ,

$$A(\theta) = \int D(\theta')\Psi(\theta - \theta') d^2\theta' \,. \tag{4.3}$$

The variance of the estimate is

$$\sigma_A^2 = \frac{1}{(2\pi)^2} \int |\hat{\Psi}(k)|^2 P_N(k) d^2k \,. \tag{4.4}$$

The filter in Equation 4.1 was constructed requiring that the estimate of the amplitude of the signal A given by Equation 4.3 is unbiased and has minimum variance.

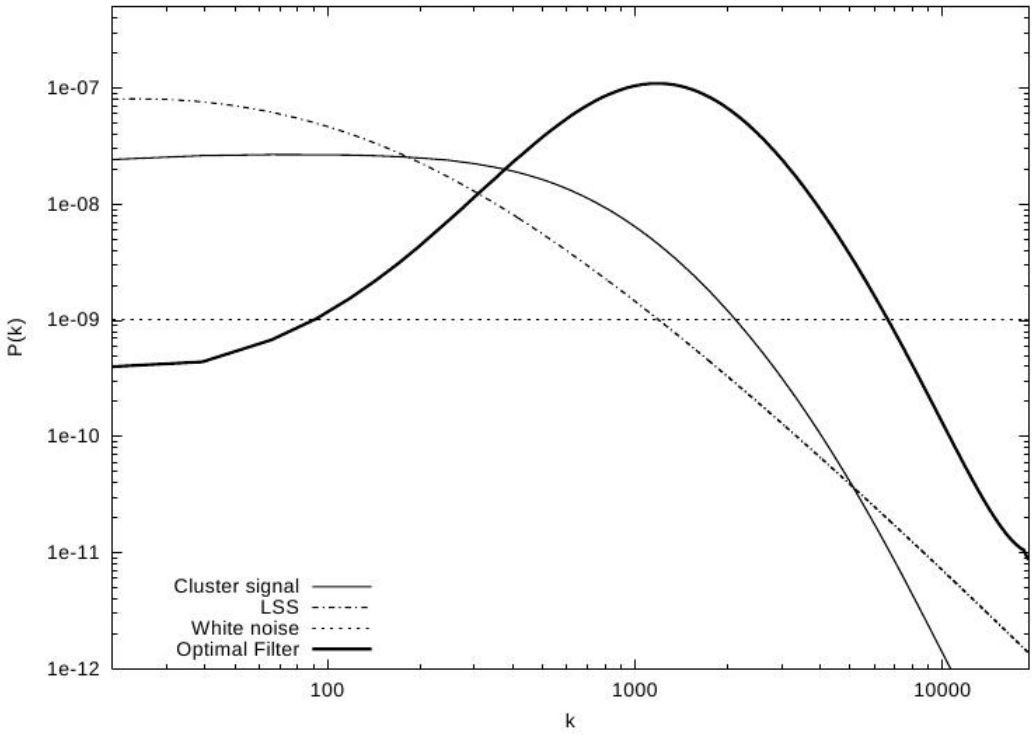

Figure 4.1: Cluster signal, noise produced by large-scale structure and by intrinsic ellipticity, and resulting shape of the filter, as a function of wavenumber k. Figure taken from Maturi et al. (2005).

The application of the weak lensing filter to the data is done evaluating $A(\theta)$ over a grid of angular positions. Equation 4.3 is then discretized to be applied to real data. It becomes

$$A(\theta) = \frac{1}{n_g} \sum_{k=1}^{N} \epsilon_{tk} \Psi(|\theta_k - \theta|) \ , \tag{4.5}$$

where n_g is the number density of galaxies used for weak lensing measurement and ϵ_{tk} denotes the tangential component of the k-th galaxy ellipticity with respect to the angular position. Thus, the filtering is performed in the real domain, which requires to back-Fourier transform the filter function defined in Equation 4.1. For every estimate of A, its error σ_A^2 is also computed, according to Equation 4.4. Peaks with S/N > 3 are selected from the map and considered as detections.

4.3 Optimal filtering: galaxy distribution

4.3.1 The filter

To derive the optimal filter, we assume that positions, magnitudes (at least in one band), and possibly photometric redshifts of the galaxies in a surveyed field are available. Furthermore, we assume that bound structures like galaxy clusters present in the field can be traced by the galaxy distributions within them and that the spatial distribution of background galaxies is random. Under these assumptions, given the characteristics of the survey, bands of observation and depth, it is possible to define a cluster model, describing the expected spatial and magnitude distributions of the cluster galaxies and a noise model describing the background distribution of field galaxies.

Thus we define a model $n_m(\theta, m)$ for the observed distribution of galaxies as a function of the position θ and magnitude m, given by the sum of a field and a cluster component (Postman et al., 1996):

$$n_m(\theta, m) = n_f(m) + \Lambda n_c(\theta, m) = n_f(m) + \Lambda P(\theta - \theta_c)\phi(m) \ , \qquad (4.6)$$

where $n_f(m)$ is the magnitude distribution of field galaxies, θ_c is the cluster centre, P is the projected radial profile of the cluster galaxies number density, $\phi(m)$ is the cluster luminosity function and Λ is a richness parameter, proportional to the total number of cluster galaxies.

The detection of a cluster in the input catalogue is based on the comparison of the observed distribution of galaxies to the model n_m. More specifically, for a fixed cluster centre θ_c, we compute the value of Λ for which our model distribution best describes the observed distribution $n_d(\theta, m)$. The likelihood of the observed data n_d given the model n_m is

$$\begin{aligned} \mathcal{L} &= -\int \frac{[n_d(\theta, m) - n_m(\theta, m)]^2}{n_m(\theta, m)} d\Omega dm \\ &= -\int \frac{[n_d(\theta, m) - n_f(m) - \Lambda n_c(\theta, m)]^2}{n_f(m)} d\Omega dm \ , \end{aligned} \qquad (4.7)$$

assuming that the background noise is Poissonian noise and that the galaxy density is high enough such that the Gaussian approximation holds.

We derive the value of the richness Λ that maximizes $\mathcal{L}$ by imposing

$$\frac{d\mathcal{L}}{d\Lambda} = 2\int \frac{n_c}{n_f}(n_d - n_f) d\Omega dm - 2\Lambda \int \frac{n_c^2}{n_f} d\Omega dm = 0 \ , \qquad (4.8)$$

which leads to

$$\begin{aligned}\Lambda &= \frac{\int \frac{n_c}{n_f}(n_d - n_f)d\Omega dm}{\int \frac{n_c^2}{n_f}d\Omega dm} \\ &= \frac{\int \frac{n_c}{n_f}n_d d\Omega dm - \int n_c d\Omega dm}{\int \frac{n_c^2}{n_f}d\Omega dm} \,. \end{aligned} \tag{4.9}$$

Only the first term in the numerator depends on the spatial distribution of the data with respect to θ_c. Thus, Equation 4.9 can be re-written as

$$\Lambda = \int \Phi(\theta - \theta_c, m) n_d(\theta, m) d\Omega dm - B \,, \tag{4.10}$$

where Φ is the optical linear filter defined as

$$\Phi(\theta - \theta_c, m) = \left(\int \frac{n_c^2}{n_f}d\Omega dm\right)^{-1} \frac{n_c(\theta, m)}{n_f(m)} \,, \tag{4.11}$$

and B is the contribution of background galaxies that is subtracted to work on a zero-mean noise field,

$$B = \left(\int \frac{n_c^2}{n_f}d\Omega dm\right)^{-1} \int n_c d\Omega dm \,. \tag{4.12}$$

If we insert Λ obtained from Equation 4.9 back into Equation 4.7, we obtain the corresponding value of the likelihood

$$\mathcal{L} = \mathcal{L}_0 + \frac{\left[\int \frac{n_c}{n_f}(n_d - n_f)d\Omega dm\right]^2}{\int \frac{n_c^2}{n_f}d\Omega dm} \,, \tag{4.13}$$

where $\mathcal{L}_0$ is a (negative) constant that does not depend on the position of the cluster θ_c. Comparing Equations 4.13 and 4.9, we see that the varying part of $\mathcal{L}$ is proportional to Λ^2. This squared dependence of $\mathcal{L}$ on Λ reflects the fact that the χ^2 approach returns high likelihoods also for galaxy underdensities, for which Λ would be negative.

This is why we consider convenient to search for galaxy clusters as peaks of the distribution of Λ as a function of the sky position θ_c, instead of likelihood maxima. In fact, we take full advantage of the linear response to the data, provided by Equation 4.10, and of the physical direct interpretation of Λ, which is a suitably normalized number of visible galaxies belonging to the cluster (see Section 4.3.2).

The variance of Λ is given by $\sigma_\Lambda^2 \equiv \langle(\Lambda - \hat{\Lambda})^2\rangle$, where $\hat{\Lambda}$ is an estimate of the true value of Λ, given by Equation 4.10. Since $\Lambda\hat{n}_c(m,\theta)$ and $\hat{n}_f(m)$ are random realizations of the cluster and field populations, the resulting σ_Λ^2 has two contributions: the first term is induced by the random fluctuations in the background, the second by the random sampling of the cluster model:

$$\sigma_\Lambda^2 = \left(\int \frac{n_c^2}{n_f} d\Omega dm\right)^{-1} + \Lambda \frac{\int \frac{n_c^3}{n_f^2} d\Omega dm}{\left(\int \frac{n_c^2}{n_f} d\Omega dm\right)^2} \,. \tag{4.14}$$

By applying the formalism of linear matched filters, one can see that the filter minimizes the variance of the estimated Λ due to random fluctuations only. The second term in Equation 4.14 gives the intrinsic Poissonian fluctuation of the cluster realization.

In case multi-band observations are available and a model for the colour distribution is assumed, the previous equations can easily be generalised to use a magnitude vector $\vec{m}$, instead of a scalar m.

4.3.2 Modelling the galaxy clusters - single-band case

In the single band case, the cluster optical model is specified by the spatial and luminosity distribution of cluster galaxies. We take these distributions from the analysis of observed galaxy clusters. For simplicity, we assume spherical symmetry. We also assume that the projected radial distribution of the cluster members follows an NFW model, as the matter distribution (see Eq. 4.2). Our choice is motivated by the analysis made by Hansen et al. (2005) on a large sample of optically-selected clusters from the SDSS. They found that the NFW model is a good description of the cluster galaxy distribution up R_{200}. The concentration parameters derived from galaxy distributions described by the NFW model depends on the richness but it is

in general smaller ($1 < c < 3$) than those found in numerical simulations for dark matter halos of similar mass. In this context we use a scale radius $r_s = 500$ kpc$/h$, that represents a good estimate for quite rich clusters (see Figs. 7 and 8 in Hansen et al., 2005). The resulting radial profiles computed at different redshifts are shown in Fig. 4.2, as examples. For the cluster luminosity distribution, we assume the Schechter luminosity functions (Schechter, 1976), with parameters taken from Popesso et al. (2005), who analysed 97 clusters observed in the SDSS. Since these are calculated from galaxy counts inside a radius of 1 Mpc$/h$ from the cluster centre, we apply the same cut to the radial density profile. For a realistic concentration parameter $c = 2$, this implies that we limit our analysis within R_{200}, where the projected NFW is a good description of observed clusters, as discussed. Then, we normalize properly the cluster model n_c to set the richness parameter Λ to be the total number N of visible galaxies in a cluster. Integrating the cluster model, we get:

$$N = \Lambda \int n_c(\theta, m) d^2\theta dm \tag{4.15}$$

$$= \Lambda \iint P(\theta) d^2\theta \int \phi(m) dm \; . \tag{4.16}$$

To set $N = \Lambda$, we impose the integrals of the angular distribution and the luminosity function to be normalized to unity, that is

$$\iint P(\theta) d^2\theta = 1 \; , \tag{4.17}$$

$$\int_0^{m_{\rm lim}} \phi(m) dm = 1 \; , \tag{4.18}$$

where $m_{\rm lim}$ is the limiting magnitude of the sample.

4.3.3 Modelling the galaxy clusters - multi-band case

In this Section we describe another way to build the galaxy cluster model, that allows us to derive the galaxy content of a cluster for larger ranges of observing bands and cluster richness. In particular, this method is well-suited to derive a colour distribution for the cluster galaxies, and thus to consider combined multi-band information in the analysis. We will make use of this cluster model in Sect. 5.2. The main ingredients for this method are taken

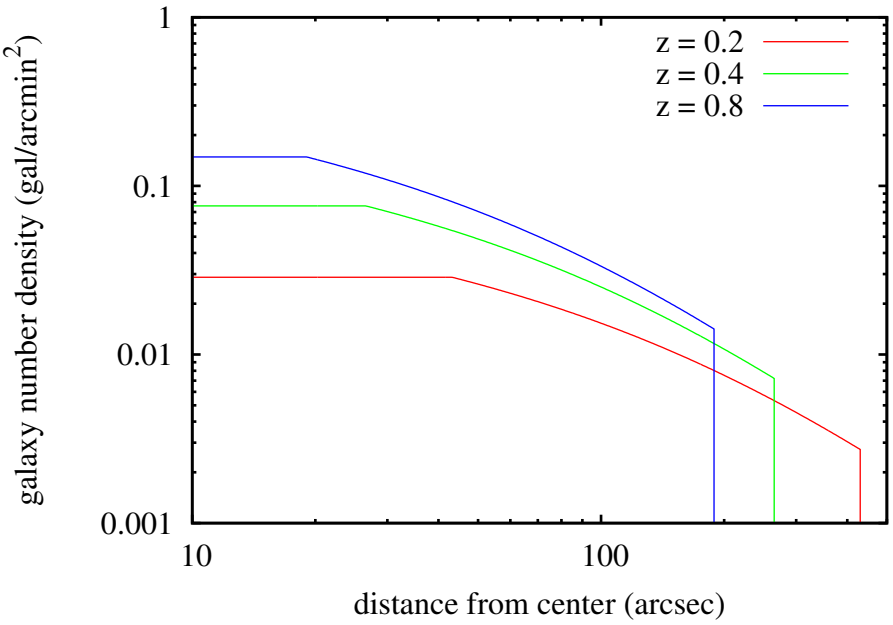

Figure 4.2: Radial profiles of the cluster model at different redshifts, as labelled in the figure. The profile is flattened inside 100 Kpc/h to avoid the central divergence, while the cut off at large radius corresponds to the R_{200} limit.

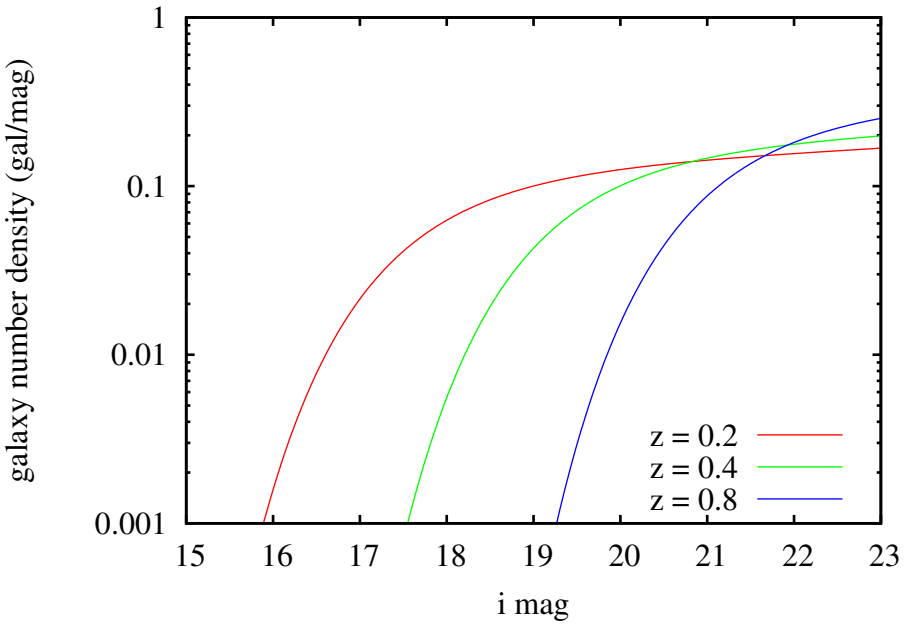

Figure 4.3: Magnitude functions of the cluster model at different redshifts, as labelled in the figure.

from the analysis of the MaxBCG cluster catalogue extracted from SDSS data (Hansen et al., 2009; Sheldon et al., 2009; Hao et al., 2009). We use the relations between M_{200}, r_{200} and N_{200} derived in Sheldon et al. (2009):

$$M_{200} = 1.7510^{12} h^{-1} M_{\odot} N_{200}^{1.25} \,, \tag{4.19}$$

$$R_{200} = 0.182 h^{-1} \mathrm{Mpc} N_{200}^{0.42} \,. \tag{4.20}$$

The angular density of galaxies as a function of angular distance r to the cluster center, in terms of R_{200}, is taken from Sheldon et al. (2009), for different values of N_{200}, as shown in Fig. 4.4. In order to derive a model for the cluster appearance in different observing bands, we need to assume a spectral-energy distribution for the cluster galaxies. In particular, we split them in two broad classes, blue and red galaxies, following the well-known bimodality of galaxy colours (see e.g. Balogh et al., 2004). For each point in the $(r, \vec{m})$ space, we write the density of galaxies as the sum of the contributions given by the red and the blue populations,

$$n(\vec{m}, r) = n_{red}(\vec{m}, r) + n_{blue}(\vec{m}, r) \,. \tag{4.21}$$

Each contribution is calculated as

$$n_i(\vec{m}, r) = f_i \phi(m_1) \prod_{j=1}^{n_b - 1} G(m_j - m_1) \,, \tag{4.22}$$

where f_i is the fraction of galaxies belonging to that type, $\phi(m_1)$ is the luminosity function in the first band and $G(m_j - m_1)$ is the colour distribution. The fraction of red galaxies f_{red} as a function of the distance from the cluster center is taken again from Hansen et al. (2009), for different values of N_{200}. The luminosity functions in the i band for the two kind of galaxies are described with Schechter functions by Hansen et al. (2009). Then we use the work by Loh et al. (2008), who compared the observed colours of the red and blue clumps to the standard templates of Coleman et al. (1980) and Kinney et al. (1996), which we already introduced in the previous Chapter. We thus express the mean SEDs of the blue and red populations as linear interpolations of the aforementioned template SEDs. In this way we calculate the colour and the k-correction and obtain the proper m^* for each kind of galaxy, from the one observed in the i band. Using the same SEDs we

calculate also the mean colour for each kind of galaxy, in case multi-band information is needed. A Brighter Cluster Galaxy (BCG) is added at the center of the cluster, with an i-band luminosity given by

$$L_{BCG} = 0.387 \log N_{200} - 0.312 \,, \tag{4.23}$$

and a SED analogue to the one for red satellites. Differently from the single-band case described in the previous Section, the normalisation of the model is done in such a way that $\Lambda = 1$ corresponds to a realistic galaxy distribution for a cluster of the chosen mass M_{200}.

Apart from the cluster model, the other ingredient of our recipe is the distribution of field galaxies, that we assume to be random with a fixed mean angular number density. The magnitude distribution of the field galaxies $n_f(m)$ can be estimated from the distribution of the whole galaxy sample, provided that the field is large enough such that the contribution of cluster galaxies can be considered negligible.

4.3.4 Map making

Clusters are searched in three-dimensional space, where their positions are given by the angular coordinates θ_c and by the redshift z_c. We create maps of Λ estimates for different values of the redshift, and we search for cluster detections in these maps. Comparison of detections in different redshift slices is done subsequently (see Section 4.3.5).

At a fixed redshift z_c, we build a two-dimensional map evaluating Λ over a grid of angular positions θ_c. Since we deal with discrete quantities, the integral in Equation 4.9 must be approximated as a sum over the galaxy positions. Only the first term in the numerator needs to be evaluated for every grid position θ_c, while the second one is equal to unity because of the normalization. The denominator for a given redshift is a constant for all sky positions. We denote it by C. Thus the discretized version of Equation 4.9 reduces to

$$\Lambda_{ij} = \Lambda(\theta_i, \theta_j) = \frac{A_{ij} - 1}{C} \,, \tag{4.24}$$

where

$$A_{ij} = \sum_{k=1}^{N} \frac{n_c(r_k, \vec{m}_k)}{n_f(\vec{m}_k)} = \sum_{k=1}^{N} \frac{P(r_k)\phi(\vec{m}_k)}{n_f(\vec{m}_k)} \,. \tag{4.25}$$

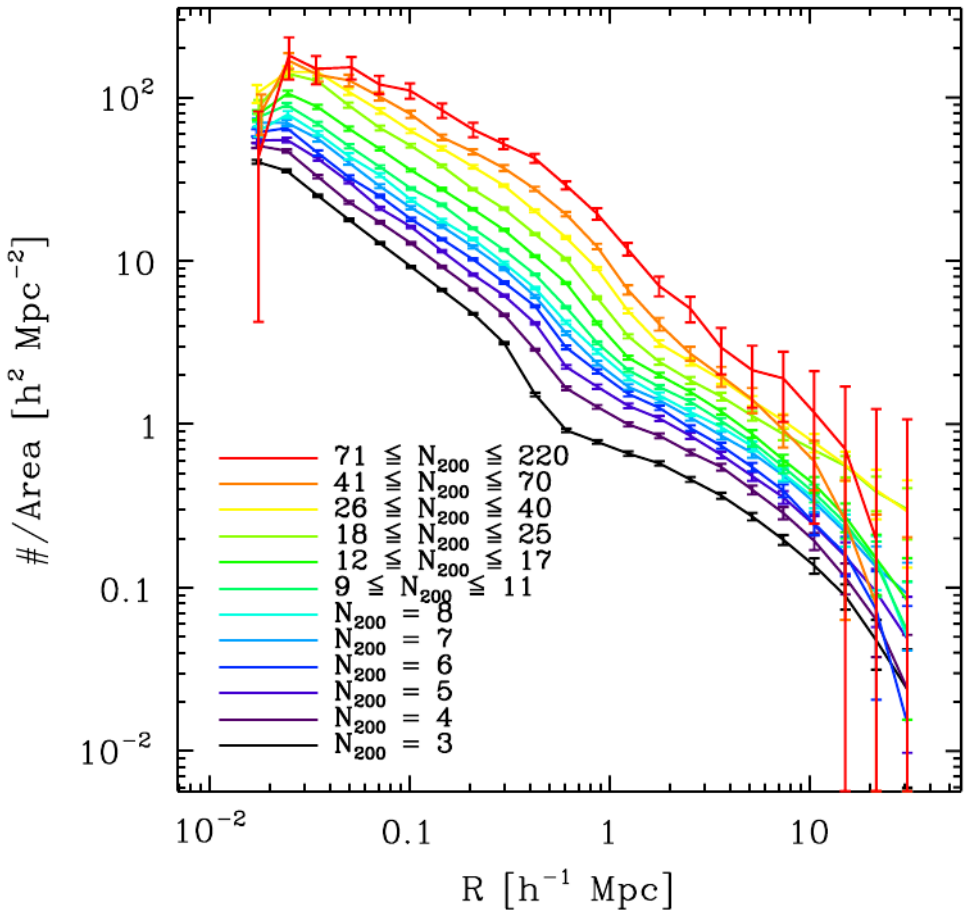

Figure 4.4: Distribution of galaxies as a function of angular distance to the cluster center. Figure taken from Sheldon et al. (2009)

In the formula above, r_k is the angular distance of the galaxy from the centre

$$r_k = |\theta_k - (\theta_i, \theta_j)| \,, \tag{4.26}$$

and the index k runs over all N galaxies of the sample. Clusters candidates are identified with the peaks of the resulting Λ-map. To select only significant peaks, we calculate their signal-to-noise ratios using the variance given in Equation 4.14.

4.3.5 Redshift determination

To estimate the cluster redshift, we cannot just maximize Λ with respect to z_c. Indeed, this would introduce a bias as discussed below. The analytic response of the algorithm as a function of the search redshift z_c for a cluster at a redshift $\bar{z}_c$ and with a galaxy distribution $\bar{n}_c$ is

$$\Lambda(z_c) = \frac{\int \frac{n_c}{n_f} \bar{n}_c d\Omega dm}{\int \frac{n_c^2}{n_f} d\Omega dm} \,. \tag{4.27}$$

This corresponds to Equation 4.9 when applied to a known distribution $n_d = n_f + \bar{n}_c$. By construction, $\Lambda = 1$ when $z_c = \bar{z}_c$, but there is no reason why this should be a maximum for the function $\Lambda(z_c)$. This is shown in Fig. 4.5, where we plot the analytic response for two clusters located at $z = 0.2$ and $z = 0.4$. For both of them the maximum of Λ as a function of z_c is found at a higher redshift than the cluster's.

For a correct redshift estimate, the likelihood $\mathcal{L}$ (Equation 4.13) in the peaks of the 2D distribution of Λ must be used. By definition, the value of $\mathcal{L}$ is maximal when the cluster is filtered by a model at the correct redshift (See Fig. 4.6).

If the photometric redshift is available for each galaxy, we can use this further information in our analysis. Assuming that a redshift probability distribution $p_k(z)$ is derived for the k-th galaxy of the sample, $p_k(z_c)$ can be used as a weight factor for that galaxy when estimating the field number density, n_f, the cluster richness, Λ, and the likelihood, $\mathcal{L}$, at a redshift z_c..

On the other hand, this can be interpreted as an extension of the two-dimensional linear matched filter to a three-dimensional one, where the redshift probability function represents the cluster profile along the line of sight.

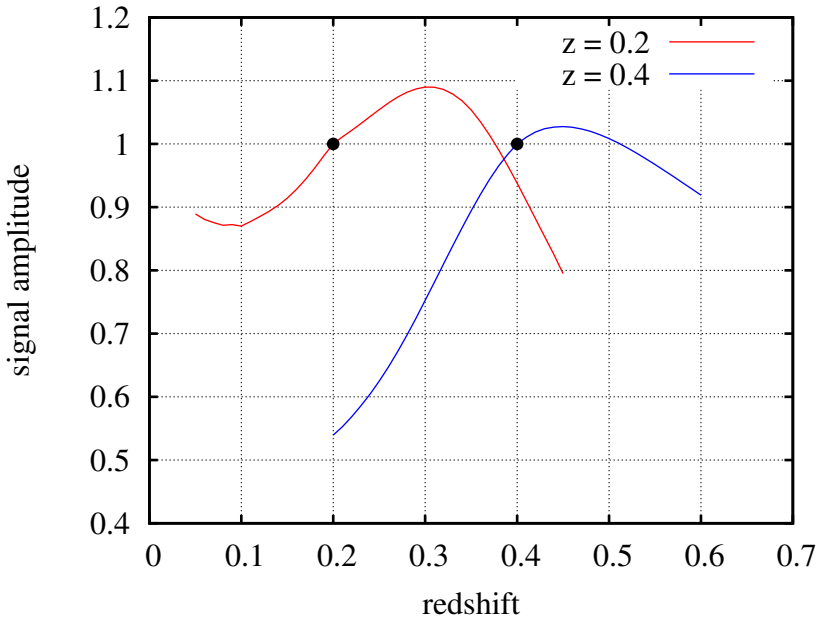

Figure 4.5: Analytic response of the algorithm (Eq. 4.27) for two clusters at redshifts 0.2 and 0.4. When the cluster is seen at the correct redshift, the value of the response is unity by definition (black points).

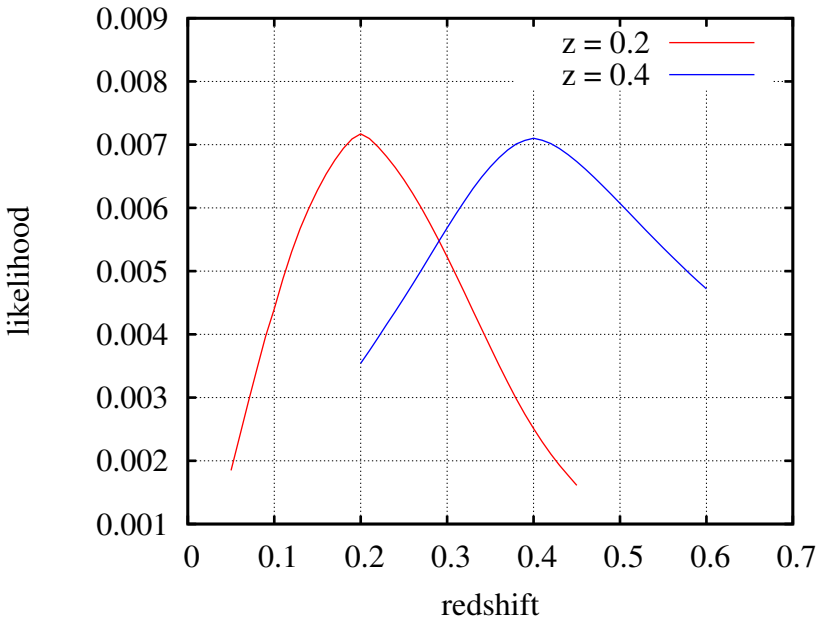

Figure 4.6: Analytic value of $\mathcal{L}$ for two clusters at redshift 0.2 and 0.4, as a function of the search redshift z_c.

In fact, in an ideal case where all galaxies have the same redshift probability distribution, which for simplicity we assume to be a Gaussian distribution centered on their most probable location, $z_{est,k}$, with a constant rms, σ_z,

$$p_k(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left(-\frac{(z_{est,k}-z)^2}{2\sigma_z^2}\right), \tag{4.28}$$

the three-dimensional model with a cluster at redshift z_c is

$$\begin{aligned} n_m(\theta,m,z) &= n_f(m,z) + \Lambda n_c(\theta,m,z) \\ &= n_f(m,z) + \Lambda P(\theta-\theta_c)\phi_z(m)p(z)\ , \end{aligned} \tag{4.29}$$

where

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left(-\frac{(z_c-z)^2}{2\sigma_z^2}\right). \tag{4.30}$$

If we then further assume that the background distribution is only slowly changing with redshift, so that the background $n_f(m,z)$ for $z_c - \sigma_z < z < z_c + \sigma_z$ is constant, the richness estimate at redshift z_c is

$$\Lambda = \frac{\displaystyle\int \frac{n_c(\theta,m,z)}{n_f(m,z_c)}(n_d(\theta,m,z) - n_f(m,z_c))d\Omega dm dz}{\displaystyle\int \frac{n_c^2(\theta,m,z)}{n_f(m,z_c)}d\Omega dm dz}\ . \tag{4.31}$$

This is equivalent to Equation 4.9, but with the additional dimension. In practical applications, Equation 4.31 has to be discretized in analogy with Equation 4.24. Thus,

$$A_{ij} = \sum_{k=1}^{N} \frac{P(r_k)\phi(m_k)p_k(z_c)}{n_f(m_k,z_c)}\ . \tag{4.32}$$

In this case the numerator is evaluated weighting every galaxy with $p_k(z_c)$. The constant denominator C, that does not depend on the observed galaxy spatial distribution, can be evaluated using a model for σ_z, possibly depending on the magnitude and the redshift estimate of the galaxies. The same model is also necessary for the evaluation of the variance that, in analogy with Equation 4.14, is

$$\sigma_\Lambda^2 = \left(\int \frac{n_c^2}{n_f}d\Omega dm dz\right)^{-1} + \Lambda \frac{\displaystyle\int \frac{n_c^3}{n_f^2}d\Omega dm dz}{\left(\displaystyle\int \frac{n_c^2}{n_f}d\Omega dm dz\right)^2}\ . \tag{4.33}$$

The filter is applied to the data as usual, creating maps for different redshift slices and selecting peaks, as described in Section (4.3.4). The most likely redshift estimate for a detection is then chosen.

4.3.6 Comparison with other optical cluster finding methods

The optimal matched filter we use for cluster detections (both optical and weak lensing) relies on the general knowledge we have of galaxy clusters, as incorporated in the cluster model, and selects all physical properties which enables us to distinguish them from the field. It is in fact a general approach which can incorporate at once different cluster aspects in contrast to other methods which aim at a single signature such as galaxy overdensities, cluster galaxy red-sequence, cD galaxies, etc. In other words, instead of enforcing some criteria which is later applied to data, we first define the general properties of galaxy clusters and then let the filter find what are the distinctive features for their detection given the actual data noise.

In the rest of this Section we discuss the main techniques applied in literature and how they compare with the one presented in this work.

Looking for galaxy overdensities in astronomical images by eye was the first method used to detect galaxy clusters. In the last decades different automated methods were proposed, such as friend-of-friend algorithms (Li & Yee, 2008), Voronoi tesselation (Ramella et al., 2001), filtering by wavelets (Finoguenov et al., 2007), adapting kernels (Mazure et al., 2007) or matched filters (e.g. Postman et al., 1996; Kepner et al., 1999; Gilbank et al., 2004; Dong et al., 2008; Menanteau et al., 2009; Milkeraitis et al., 2009). Together with the additional features discussed in the following sections, our algorithm incorporates the mentioned matched filters. In fact it includes the angular clustering information on the sky according to the P term of Eq. 4.32 or, if photometric redshifts are available, the full three-dimensional clustering information thanks to the p term of Equ. 4.32. In more details, the main differences with respect to the matched filters mentioned in this Section are:

1. our filter can take advantage of the information from photometric redshifts in a very flexible way, accounting for their uncertainty. Instead of simply slicing the data-set and neglecting all galaxies outside a fixed

redshift range (e.g. Milkeraitis et al., 2009), we consider all galaxies belonging to the data set by weighing them according to their own redshift probability distribution.

2. It uses distinct luminosity functions for galaxy clusters and field galaxies to define the cluster model and the noise contribution. Although this seems an obvious choice, previous filtering algorithms use the luminosity functions taken from the whole population of galaxies, including the field, to define the cluster model.

3. It is linear with respect to the data, because it makes use of the richness Λ instead of the likelihood $\mathcal{L}$ as the main object of analysis. In fact, even if they are based on the same statistical distribution, Λ is linear, in contrast with $\mathcal{L}$ which is quadratic, and therefore is capable to distinguish positive overdensities from negative underdensities to which $\mathcal{L}$ would assign high probabilities.

4. It gives as a natural output an estimate of the number of observed galaxies in a cluster, which can be easily corrected for redshift dimming to get a meaningful physical quantity (see Section 4.4.3). With the second implementation of the cluster model, we also have a measure of the galaxy population of the cluster in terms of the expected population for a cluster of mass M_{200}.

The algorithm used by Koester et al. (2007), based mainly on red-sequence information, includes a cut-off term for the brighter cluster galaxy luminosity, such that detections without a very brilliant central galaxy are discarded. In our approach, instead of looking for a specific cluster member, we let the filter select in an optimal way the whole luminosity function ϕ expected for a cluster at a given redshift in contrast with the field luminosity function n_f, as in Equation 4.32. As the most massive and luminous galaxies are found preferably at the center of the clusters, at low magnitudes the ratio between the cluster model luminosity function and the observed field distribution will increase. The algorithm will assign large weights to very luminous galaxies and thus they will be a strong indication of the presence of clusters. The advantage of our approach is that the filter is defined according to well defined statistical quantities and not to a free parameter. In addition we allow the

algorithm to be sensitive to a broad range of systems, from rich clusters to groups which might otherwise fall below a given threshold.

Red-sequence cluster finders rely on the fact that the center of galaxy clusters is typically populated by elliptical galaxies whose color is on average redder than the one of the field galaxies at the same redshift. Usually a cut in a colour-magnitude diagram is performed to select these red galaxies overdensities that identify the galaxy cluster candidates. This technique, presented in Gladders & Yee (2000), has been succesfully used in different galaxy surveys (e.g. Gladders & Yee, 2005; Koester et al., 2007; Lu et al., 2009; Thanjavur et al., 2009). When adopting the second method for model creation, we make use of the colour information by defining a multi-band luminosity distribution for the cluster. The expected density in the magnitude-colour space depends explicitly on the fraction of red objects in galaxy clusters. As this is known to be larger than the corresponding field value, red galaxies will receive large weighs by the algorithm, and thus will be markers of the presence of a cluster. As in the case of the BCG, we underline that this is done not by defining a cut-off in the color space but by defining the proper weight to each galaxy performing an optimized analysis.

4.4 Application to the COSMOS field

In this Section, we will show the application of the algorithm to real data, namely the i' band galaxy catalogue of the COSMOS field (Ilbert et al., 2009), limited at $i' = 25$. In this work, the first method to draw the galaxy cluster model is used. Before applying the filter to data, we have to evaluate the capability of our algorithm to detect clusters and to correctly measure their richness and redshift. In these tests we use the field distributions measured from the COSMOS data to define the filter. Thus, we use the specific implementation of the algorithm that will be later adopted for the final data analysis.

4.4.1 Signal-to-noise estimates

We now evaluate the expected signal-to-noise ratios for clusters with different redshift and richness, as described by our cluster model, once the actual prop-

erties of the galaxies in the COSMOS catalogue are assumed. The expected signal-to-noise ratios when the photometric redshifts are neglected are shown in the upper panel of Fig. 4.7. In this case, the expected signal and noise are given by Equations 4.9 and 4.14, respectively, and the signal-to-noise ratio peaks at $z \sim 0.45$. To explain this behaviour, we plot separately in Fig. 4.8 the expected signal of a cluster with 100 visible galaxies at $z = 0.2$ and the two noise terms as a function of redshift. The signal, i.e. the number of visible galaxies of the modeled cluster, decreases for increasing z as expected (solid line). The trend for the noise is different: on one hand, the term related to the background galaxies only (dot-dashed line) grows monotonically as a function of redshift up to $z \sim 0.5$ after which it basically stays constant; on the other hand, the intrinsic fluctuations of the cluster signal decreases monotonically with z. As a result the latter term is dominant at low redshifts (up to $z \lesssim 0.4$).

This noise behaviour derives from the fact that for low-redshift clusters, the filter, being proportional to $n_c(m)/n_f(m)$, selects the bright end of the cluster luminosity function, that has almost no background. Therefore the background noise decreases with redshift while the intrinsic fluctuations of the cluster signal increase because the Λ estimate depends in practice on the very few bright galaxies only.

In order to consider the case in which photometric redshift information is available, we use the values for the redshift (z_{est}) and its 68% confidence levels (z_{min}, z_{max}) reported in the COSMOS catalogue to associate to each galaxy a probability distribution given by

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(- \frac{(z - z_{est})^2}{2\sigma^2} \right) , \tag{4.34}$$

where

$$\sigma = \begin{cases} z_{est} - z_{min} & \text{if } z < z_{est} \\ z_{max} - z_{est} & \text{if } z > z_{est} \end{cases} , \tag{4.35}$$

and

$$\sigma_m = (z_{max} - z_{min})/2 \, . \tag{4.36}$$

When estimating the normalization constants, i.e. the denominator of Equation 4.31 and the variance (Equation 4.33), which do not depend on the observed spatial galaxy distribution, we use the mean redshift error for different classes of objects, estimated from the comparison with a spectroscopic

subsample (Ilbert et al., 2009). For our data-set, these errors are:

$$\sigma_{\Delta z/1+z} = \begin{cases} 0.007 & \text{for } 0.2 < z < 1.5, \quad i' < 22.5 \\ 0.011 & \text{for } 0.2 < z < 1.5, \quad 22.5 < i' < 24 \\ 0.053 & \text{for } 0.2 < z < 1.5, \quad 24 < i' < 25 \\ 0.06 & \text{for } z > 1.5 \end{cases} . \tag{4.37}$$

As we can see in the lower panel of Fig. 4.7, using the photo-z information, the picture becomes more confused, as the sensitivity depends on the field population density $n_f(m, z)$ at the search redshift (see Equation 4.33). The only clear trend is the decrement of the sensitivity at high redshift due to the dimming of the cluster signal. In this case we do not have any worsening of the performance at low redshift because the weight given by the filter through the $n_c(m)/n_f(m, z)$ ratio selects almost the same part of the luminosity function at each redshift.

4.4.2 Simulations with mock catalogues

We then test the application of our algorithm to mock galaxy catalogues, built from the COSMOS data. To create realistic catalogues for the field galaxies, we first randomize the positions of the galaxies to cancel any structures contained in the field. We assume that the cluster galaxies do not affect the photometric properties of the overall sample because their number is negligible with respect to that of the field galaxies. On the top of this random background, we add some mock clusters. The cluster galaxies are distributed according to the model for the spatial and luminosity distribution of the cluster members (see Section 4.3.2). The photometric redshifts are assigned following a Gaussian redshift probability distribution with errors given by Equation 4.37. For illustrative purposes, we first generate a square field with side 1 degree, containing 25 galaxy clusters of different richness and redshift placed on a regular grid. In Fig. 4.9 we show the results of the analysis of this field without the usage of photometric redshifts, at a search redshift $z_c = 0.4$. All clusters leave an imprint on the Λ-map, whose strength depends on the richness and the redshift, with clusters at $z = z_c$ being brighter than the others with the same Λ.

Other simulations are done to verify the linear response of the filter with respect to the cluster richness. In this case, we assume for simplicity all clus-

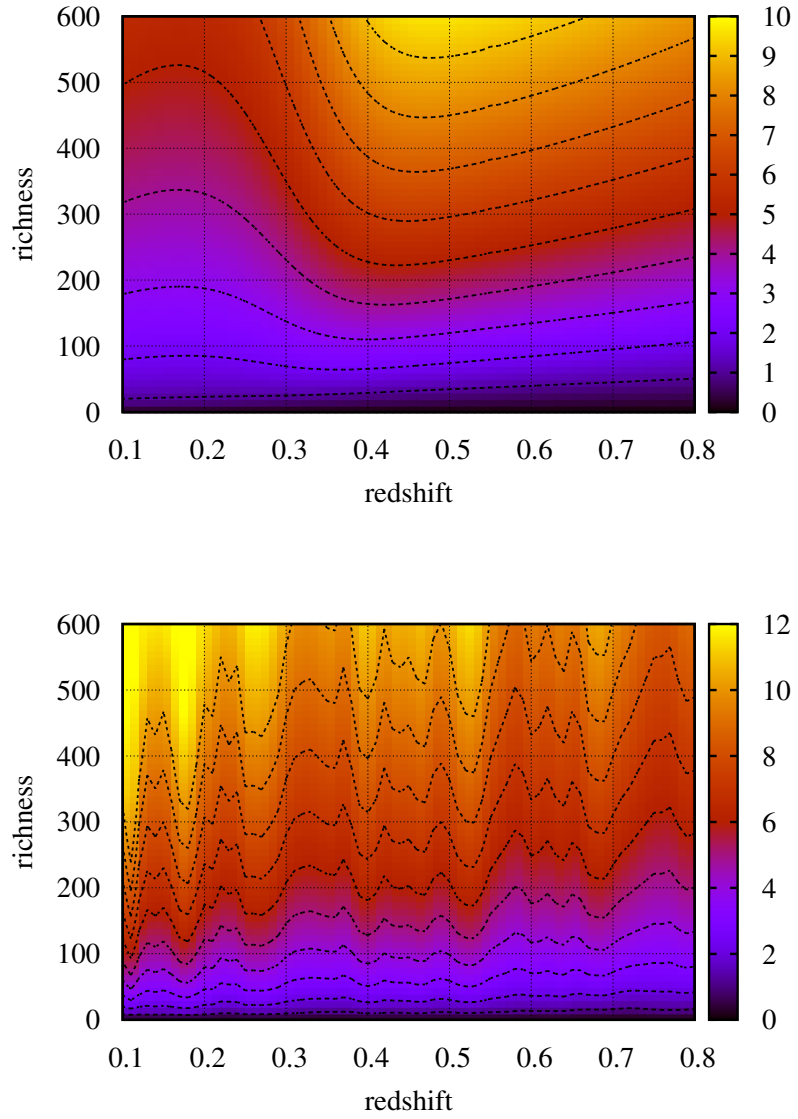

Figure 4.7: Expected signal-to-noise values for the detection of a cluster, without (top panel) and with (bottom panel) photometric redshifts, for the COSMOS i' band catalogue. The black contours refer to different integer values for S/N, starting from unity at the bottom (see the color bar on the right). The value on the y-axis is the number of galaxies of the cluster that are detected when it is located at $z = 0.2$.

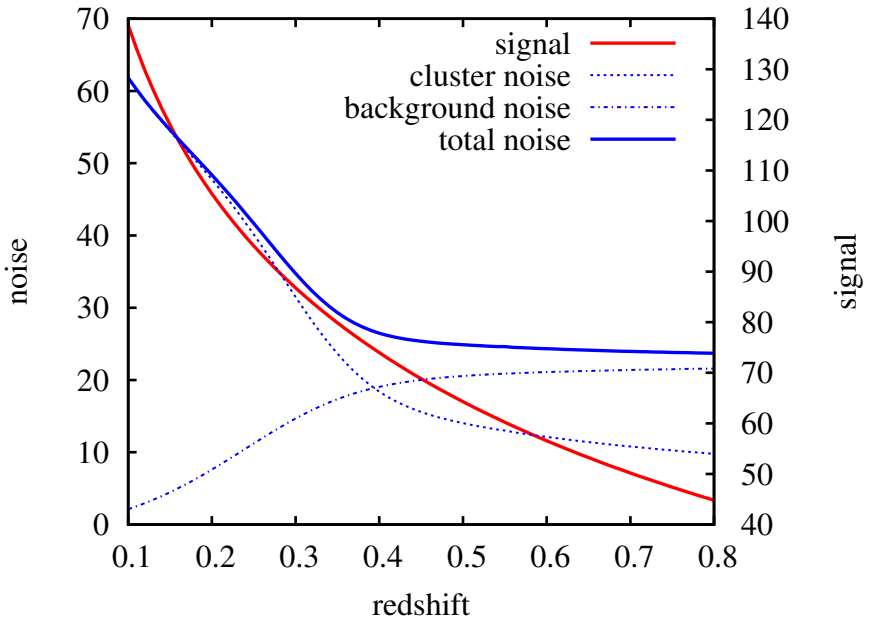

Figure 4.8: Signal and noise for a cluster with 100 visible galaxies at $z = 0.2$ as a function of redshift. The signal (in red) corresponds to the number of galaxies below $i' = 25$, calculated from the model luminosity function. The dot-dashed line represents the noise due to the background galaxies, while the dashed line represents the noise due to the fluctuations in the cluster galaxy distribution. The solid blue line represents the total noise. Scales for noise and signal are shown on the left and right, respectively.

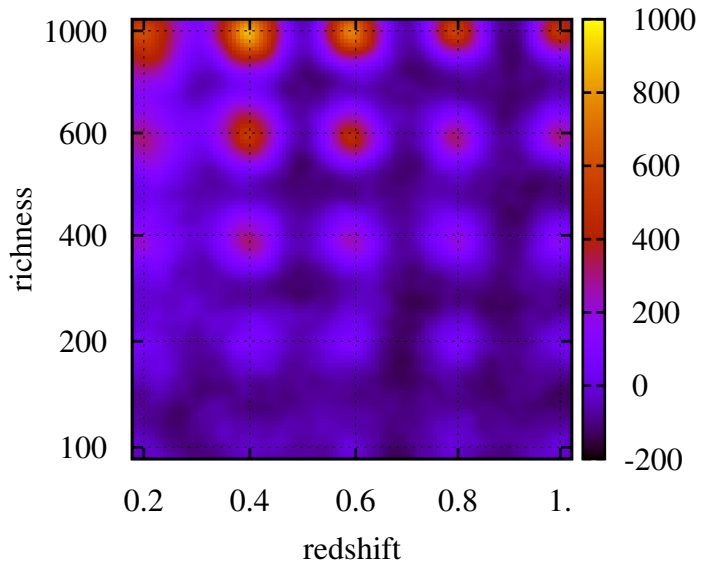

Figure 4.9: Distribution of the richness Λ at $z_c = 0.4$. Colorbar for Λ is reported on the right. The grid identifies the input positions of the mock clusters. The labels along the axes indicate the richness and redshift of the cluster at that position.

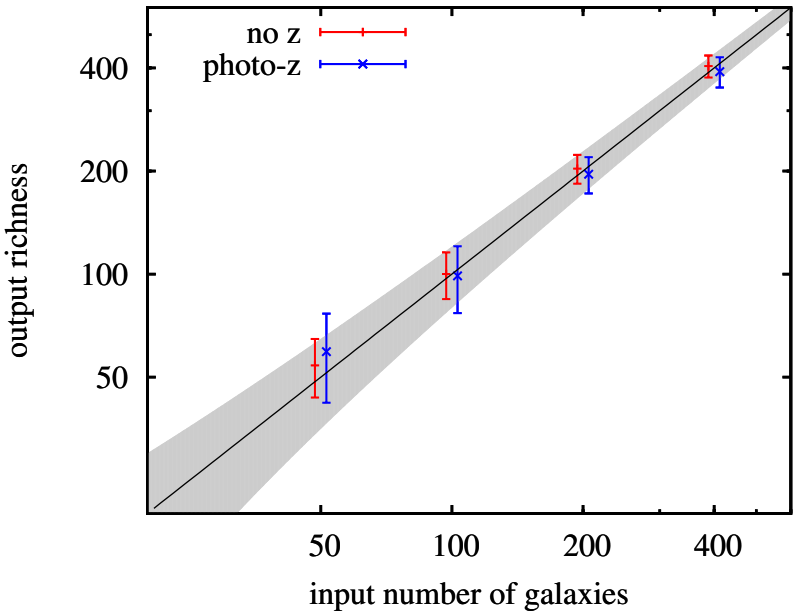

Figure 4.10: Calibration of the values for the richness Λ values obtained by the algorithm with and without redshift information. Clusters are located at $z = 0.4$. The shaded area represents the analytic estimate of the uncertainty, when the photometric redshift information is used.

ters at $z = 0.4$ and compare the input number of galaxies to the estimated one. The results are shown in Fig. 4.10 where we also verify the good agreement of the analytic estimate of the variance given in Equation 4.33 with the one resulting from our Monte Carlo simulations.

We also use simulations to test the capability of the algorithm to estimate correctly the redshift of a cluster. The results are shown in Fig. 4.11. The mock clusters were composed requiring that they have 200 galaxies under the magnitude limit at $z = 0.2$. At higher redshifts, fewer galaxies are seen and analysed, explaining the increasing variance of the measurement. Note that even with single-band observations a redshift estimate for the clusters is possible albeit with larger errors with respect to the case in which the photometric redshift information is included. A similar result was obtained by Dietrich et al. (2007).

We now apply the optical filter and the weak lensing filter to the COS-

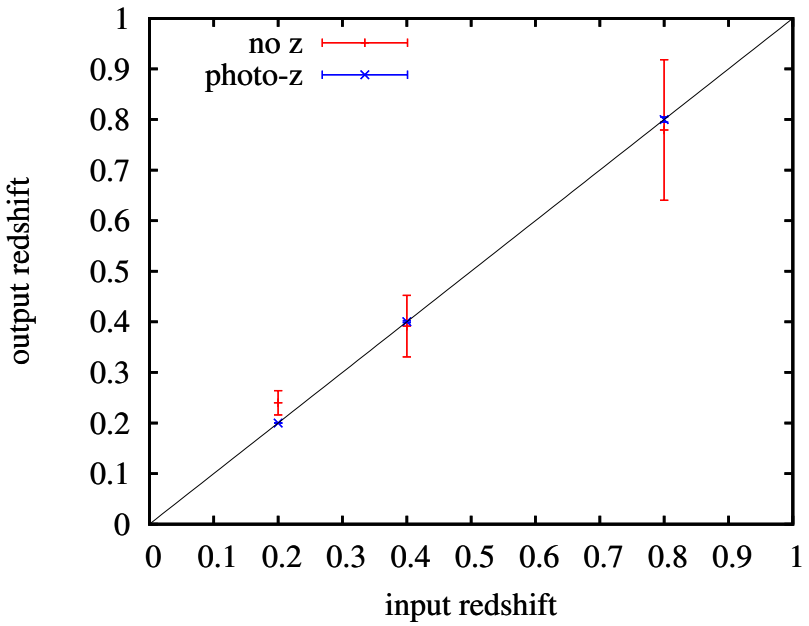

Figure 4.11: Redshift determination obtained by the algorithm, using the likelihood $\mathcal{L}$ instead of the richness Λ, with and without redshift information. When photometric redshift are used, error bars become negligible.

MOS field, a 2 square degree equatorial field that has been observed with the Hubble Space Telescope and with many other instruments, covering wavelengths from X-ray to radio (Scoville et al., 2007a).

4.4.3 Optical detections

For the optical analysis, we use the public galaxy catalogue with photometric redshifts (Ilbert et al., 2009), considering the angular position, the photometric redshift, and the i'-band magnitude measured with Subaru (Taniguchi et al., 2007). The photometric redshifts presented in the catalogue were obtained from the galaxy fluxes in 30 different bands, ranging from the ultraviolet to the near infrared. To each galaxy a SED has been assigned during the determination of the photometric redshift. We use this additional information to calculate the k-correction. We search for clusters between $z = 0.1$ and $z = 0.8$, with steps of $\Delta z = 0.02$. The upper limit of $z = 0.8$ is motivated by two factors: first, the model that we assume for the cluster luminosity function is based on low-redshift objects, and cannot be extrapolated to too high redshifts; second, if the k-correction becomes too strong for a given type of galaxies, our sample can no longer be assumed complete up to a given rest-frame magnitude. We scan an area smaller than the full field of view, such that, for each redshift, the filter (cut at $r = 1$ Mpc$/h$) is completely inside the survey field. In this way we avoid any border effect, at the price of loosing some possible detections of structures that have their centre inside the COSMOS area, but which are not completely contained in the field of view.

In Fig. 4.12, as examples, we show the results for the slices at $z = 0.48$, 0.50, 0.52 where the detections with S/N > 3 are marked with circles. Note that some structures are visible in more than one slice. In that case we assign as redshift of the detection the one where the likelihood is maximal, as discussed is Section 4.3.5. In this way we obtain 140 significant detections with a redshift estimate.

Since the absolute magnitude limit changes with redshift, the richness of clusters with the same galactic population is redshift dependent. In particular the richness decreases as the redshift increases. To correct for this effect and to obtain a quantity which depends on the cluster galaxy content only, we make use of the luminosity function $\phi(m)$ of a cluster at an arbitrary redshift $z_c = 0.2$ to compute the normalization factor

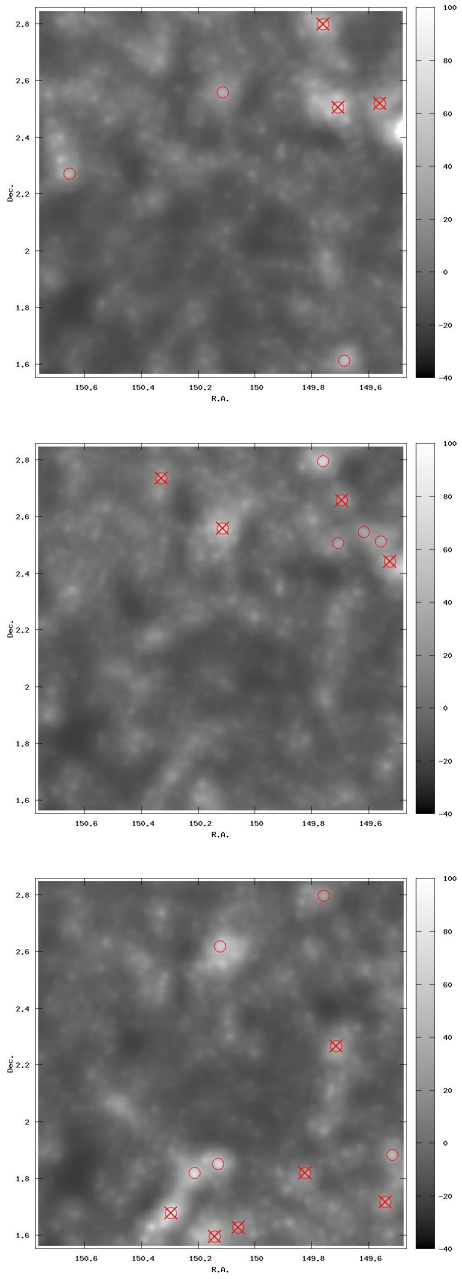

Figure 4.12: Maps of the estimates of the richness Λ on the COSMOS field, at redshifts $z = 0.48, 0.5, 0.52$, from top to bottom. Red circles indicate significant peaks, while red crosses represent a cluster detection at *that* redshift.

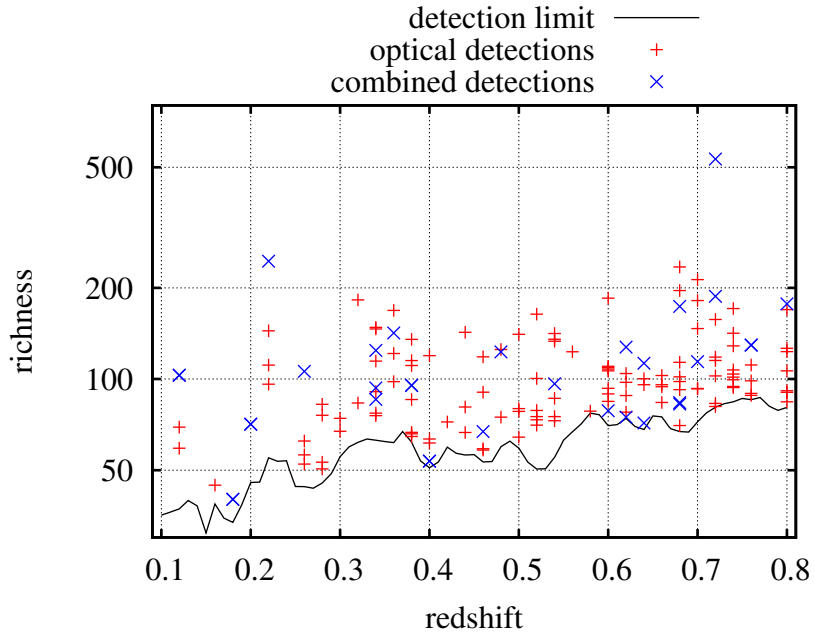

Figure 4.13: Redshift and (corrected) richness of the detected clusters. The black line represents the detection limit, that depends on model and field distributions at different redshifts. Clusters found in the optical analysis only are marked in red, those with a counterpart in the lensing analysis are marked in blue.

$$R(z) = \int_0^{m_{lim}(z)} \phi(m) dm \ . \tag{4.38}$$

This quantity represents the part of $\phi(m)$ visible at a given redshift z. Dividing the measured richness Λ for $R(z)$, where z is the estimated redshift of the cluster, we obtain a value that is directly proportional to the physical galactic content of a cluster. More precisely this 'corrected richness' corresponds to the number of cluster galaxies that would be visible if the cluster was located at $z = 0.2$. The 'corrected richness' of our detections is shown in Fig. 4.13, together with the selection threshold S/N = 3 for optical detections already derived and shown in the lower panel of Fig. 4.7. Notice that the hills and wells of the curve are due to the change of the field population as a function of redshift.

4.4.4 Lensing detections

The weak lensing cluster detections in COSMOS are performed by applying Equation 4.5 as discussed by Maturi et al. (2005) on the shear catalogue obtained with the Suprime-Cam mounted on SUBARU (Miyazaki et al., 2007). The galaxy number density in the catalogue is 42 arcmin^{-2}. Following Miyazaki et al. (2007), we assume that the mean redshift of the sources is 0.8. The weak lensing cluster model used in the filter construction is the signal expected from a cluster of mass $5 \times 10^{14} M_{\odot}$ at $z = 0.5$.

The total number of significant (e.g. with S/N > 3) detections is 82. The expected number of spurious detections at this S/N level is around 40%, as computed by Pace et al. (2007). The derived weak lensing detections are related to the optical overdensities using a correlation length of 500 kpc$/h$. This quite severe constraint is justified by the absence of any redshift information in the lensing data. The 27 detections satisfying this correlation criterion are listed in Table 4.1, while optical detections without a lensing counterpart are listed in Table 4.2.

In Fig. 4.14 we show the redshift distribution of our combined detections. From the source redshift distribution it is possible to evaluate the redshift sensitivity function for weak lensing, that is a measurement of how much the shear of observed galaxies is sensitive to structures at different redshifts. In Fig. 4.14 we also show the optical sensitivity function, computed as the S/N of a cluster with 100 visible galaxies at $z = 0.2$. The ratio of lensing-confirmed clusters over the total number of optical detections in our analysis does not show any clear trend with redshift, but this could be due to the poor statistics.

4.4.5 Comparison with literature

We can compare our results to published catalogues of clusters previously extracted from the COSMOS field. Previous analyses were done by Finoguenov et al. (2007), combining optical and X-ray information and by Olsen et al. (2007), who found galaxy overdensities, by applying a Postman-like filter to a previous version of the galaxy catalogue with no redshift information. Scoville et al. (2007b) analysed the COSMOS field with a wavelet method, and presented a catalogue of large-scale structures. We note that these structures

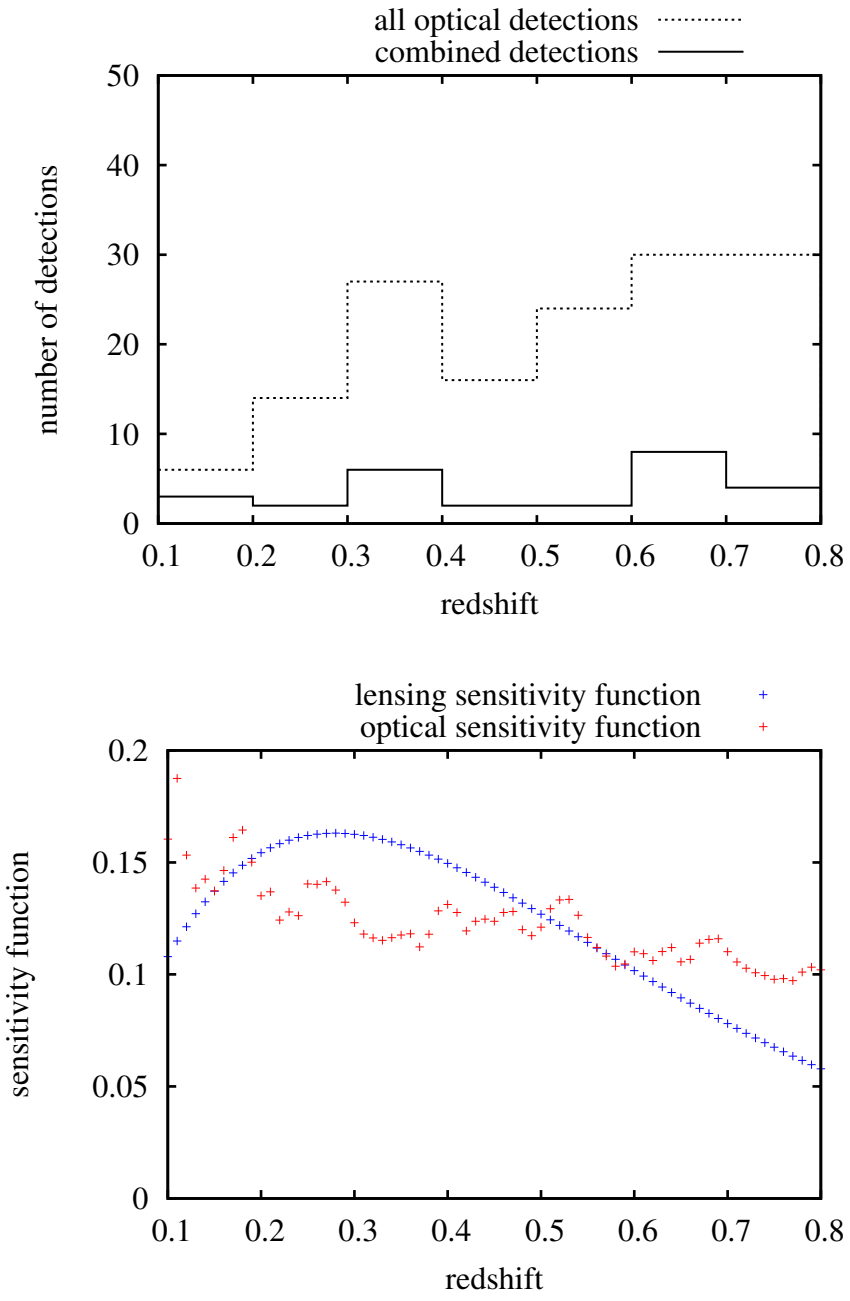

Figure 4.14: Top panel: redshift distribution of all optical detections (dashed line) and of the lensing-confirmed clusters (solid line). Bottom panel: lensing sensitivity function and optical sensitivity function, defined as the S/N of a cluster with 100 visible galaxies at $z = 0.2$, in arbitrary units.

are in general much larger than our targets, and that their size in some cases exceeds 10 Mpc in comoving coordinates. The few previous cluster detections in this field via lensing were reported by Kasliwal et al. (2008), while comparing the performance of ground-based and space-based telescopes for weak lensing, and by Miyazaki et al. (2007).

Correlating our detections with results from literature, we use different criteria depending on the characteristics of the catalogues. For sources presented in Finoguenov et al. (2007) and Olsen et al. (2007), we match detections using a physical correlation length of 1 Mpc/h, i.e. the cut-off applied to our radial filter. Instead, for the larger structures presented in Scoville et al. (2007b), we use a correlation length equal to the FWHM of their density peaks. In all these cases, the correlation along the line-of-sight is performed considering differences in redshift smaller than 0.05. This can be quite severe for Olsen et al. (2007), as their redshift determination was based on single band observations without any redshift information. For the lensing detections presented by Kasliwal et al. (2008) and Miyazaki et al. (2007), we use the same criterion we adopted for the internal matching with our lensing detections, i.e. a correlation length of 500 kpc/h. Regarding Finoguenov et al. (2007), we must stress that the comparison can only be done with their catalogue of X-ray confirmed clusters, obtained correlating 420 optical overdensities with 150 X-ray diffuse sources. This justifies some discrepancies between the two analyses.

The existence of previous detections of the objects found in our optical and lensing analysis is reported in the last columns of Tables 4.1 and 4.2.

We note here the ability of our optical filter to distinguish different galaxy clusters that are aligned along the line of sight, despite their signals are degenerate in weak lensing analyses. In fact, the two detections with identification numbers 4 and 7 correspond to SLJ1000.7+0137, a weak lensing and X-ray source studied by Hamana et al. (2009) with spectroscopic observations. They found that it is actually made by two, or possibly three, overdensities of galaxies at different redshifts, almost at the same angular position. Our method is able to disentangle the two structures at $z = 0.22$ and $z = 0.34$ using information from photometric redshifts only. A similar case is that of SLJ1001.2+0135, described by Hamana et al. (2009) as the superposition of two overdensities, located at $z = 0.22$ and $z = 0.37$. Our detection 8 corre-

ID	z	R.A.	Decl.	S/N	richness	previous detections
1	0.12	150.395	2.446	5.280	102.89	X
2	0.18	150.060	2.200	3.327	40.11	
3	0.20	150.616	2.425	3.802	70.95	X
4	0.22	150.192	1.650	6.628	244.74	XL
5	0.26	149.911	2.603	4.832	105.93	XSO
6	0.34	149.939	2.607	4.375	124.31	O
7	0.34	150.193	1.658	3.573	85.70	SL
8	0.34	150.299	1.609	3.749	93.49	XS
9	0.36	149.890	2.452	4.770	141.92	XS
10	0.38	150.142	2.053	3.831	95.29	
11	0.40	150.646	2.807	3.089	53.56	
12	0.46	150.687	2.400	3.419	67.07	
13	0.48	149.760	2.798	4.473	122.87	X
14	0.54	149.521	1.884	4.131	96.47	
15	0.60	149.926	2.519	3.202	78.57	L
16	0.62	149.572	1.882	3.018	74.84	S
17	0.62	150.590	2.473	4.035	127.47	
18	0.64	149.628	1.906	3.989	112.74	S
19	0.64	150.443	1.883	3.089	71.56	
20	0.68	149.718	1.816	3.428	82.65	
21	0.68	150.088	2.193	5.331	173.71	O
22	0.68	150.290	1.580	3.453	83.61	
23	0.70	150.308	2.406	3.975	114.17	S
24	0.72	149.921	2.521	8.721	532.42	XL
25	0.72	150.141	2.069	4.953	187.60	
26	0.76	150.641	2.804	3.764	129.46	S
27	0.80	150.437	2.763	4.705	177.08	S

Table 4.1: Catalogue of the clusters detected with both optical and lensing filters. The angular position is the one of the detection from optical data. In the last column we indicate whether the cluster has been found previously with other techniques: 'X' stands for Finoguenov et al. (2007), 'S' for Scoville et al. (2007b), 'O' for Olsen et al. (2007), 'L' for Kasliwal et al. (2008) or Miyazaki et al. (2007).

ID	z	R.A.	Decl.	S/N	richness	previous detections
28	0.12	149.862	1.765	4.279	69.35	X
29	0.12	150.432	2.630	3.925	59.13	
30	0.16	150.387	2.069	3.246	44.64	
31	0.22	150.102	2.358	5.042	144.29	XSO
32	0.22	150.395	2.460	4.397	111.20	O
33	0.22	150.620	2.707	4.070	96.12	S
34	0.26	149.842	2.682	3.436	56.19	S
35	0.26	150.089	2.474	3.640	62.42	
36	0.26	150.268	2.672	3.305	52.37	S
37	0.28	149.594	2.150	3.978	75.93	
38	0.28	149.969	2.450	3.190	50.52	S
39	0.28	150.044	2.225	4.159	82.51	SO
40	0.28	150.194	1.747	3.280	53.16	S
41	0.30	150.186	2.794	3.517	74.18	
42	0.30	150.572	1.942	3.337	67.24	X
43	0.32	150.456	2.049	5.332	182.39	S
44	0.32	150.652	2.313	3.531	83.32	S
45	0.34	149.596	2.820	4.773	146.29	XS
46	0.34	149.767	2.329	3.336	75.73	S
47	0.34	150.070	2.378	3.372	77.19	X
48	0.34	150.185	1.765	4.804	148.11	XS
49	0.34	150.373	2.444	3.691	90.87	XS
50	0.34	150.536	2.730	4.186	114.54	S
51	0.36	149.787	2.167	3.900	98.05	S
52	0.36	150.119	2.689	5.227	168.45	X
53	0.36	150.523	2.570	4.385	121.46	
54	0.38	149.675	2.413	3.133	66.70	S
55	0.38	149.767	1.625	3.112	65.94	
56	0.38	149.821	2.275	4.637	135.17	SO
57	0.38	149.966	1.678	4.250	115.08	X
58	0.38	150.234	2.474	3.609	85.62	
59	0.38	150.387	2.413	4.162	110.80	X
60	0.38	150.663	2.559	3.077	64.64	
61	0.40	149.578	2.533	3.343	61.52	

ID	z	R.A.	Decl.	S/N	richness	previous detections
62	0.40	149.971	2.748	3.401	63.40	
63	0.40	150.356	2.651	4.816	119.47	
64	0.42	149.674	2.731	3.336	72.20	
65	0.44	149.961	2.210	3.315	66.64	XO
66	0.44	150.494	2.070	5.034	142.59	XSO
67	0.44	150.690	2.021	3.695	80.85	
68	0.46	149.667	1.625	4.658	118.42	S
69	0.46	150.354	2.741	3.178	58.86	
70	0.46	150.653	2.271	4.031	90.50	
71	0.46	150.721	2.244	3.160	58.26	
72	0.48	149.561	2.519	3.407	74.82	
73	0.48	149.707	2.506	4.519	125.21	S
74	0.50	149.525	2.441	3.519	78.18	S
75	0.50	149.695	2.657	3.148	64.27	
76	0.50	150.113	2.559	4.855	140.51	XS
77	0.50	150.328	2.735	3.562	79.87	S
78	0.52	149.542	1.719	3.733	73.37	
79	0.52	149.714	2.267	3.885	78.72	O
80	0.52	149.823	1.821	3.785	75.19	
81	0.52	150.059	1.629	3.646	70.42	
82	0.52	150.141	1.597	4.453	100.43	
83	0.52	150.295	1.680	5.809	163.77	
84	0.54	150.135	1.853	4.970	135.34	
85	0.54	150.216	1.821	4.928	133.21	X
86	0.54	150.260	1.765	3.528	72.84	
87	0.54	150.360	1.627	3.883	86.33	
88	0.54	150.467	2.066	5.088	141.32	X
89	0.54	150.573	2.166	3.591	75.14	
90	0.56	149.525	1.760	4.173	123.00	
91	0.58	149.578	1.683	3.028	78.39	S
92	0.60	149.617	1.740	5.134	184.96	S
93	0.60	150.063	2.798	3.874	110.18	
94	0.60	150.253	2.346	3.334	84.34	
95	0.60	150.301	2.804	3.445	89.34	

ID	z	R.A.	Decl.	S/N	richness	previous detections
96	0.60	150.491	2.745	3.851	109.03	X
97	0.60	150.574	2.608	3.827	107.78	
98	0.60	150.616	2.780	3.527	93.14	
99	0.60	150.717	2.537	3.799	106.39	
100	0.62	150.052	2.321	3.082	77.73	
101	0.62	150.520	2.473	3.620	104.23	
102	0.62	150.707	2.759	3.306	88.25	
103	0.62	150.725	2.625	3.492	97.54	
104	0.64	150.633	2.715	3.738	100.25	
105	0.64	150.737	2.825	3.644	95.81	X
106	0.66	149.802	1.804	3.204	83.66	S
107	0.66	150.185	2.158	3.438	94.33	
108	0.66	150.196	2.238	3.613	102.74	
109	0.66	150.687	1.667	3.469	95.77	
110	0.68	149.673	2.277	3.670	92.24	
111	0.68	149.925	2.597	3.092	70.21	S
112	0.68	149.948	2.097	4.162	113.53	
113	0.68	150.010	2.119	3.810	98.07	
114	0.68	150.060	2.608	6.305	234.42	SO
115	0.68	150.173	2.518	5.706	195.95	S
116	0.68	150.257	1.968	3.895	101.70	
117	0.70	149.647	2.828	3.514	93.08	
118	0.70	149.964	2.673	4.606	146.79	S
119	0.70	149.986	2.578	5.200	181.54	XSO
120	0.70	150.003	2.451	5.684	212.81	S
121	0.70	150.152	2.601	3.509	92.83	S
122	0.72	149.899	2.394	3.789	118.18	S
123	0.72	150.086	2.460	4.479	157.29	S
124	0.72	150.108	2.565	3.475	102.36	S
125	0.72	150.207	2.361	3.056	83.12	S
126	0.72	150.593	2.129	3.011	81.17	
127	0.72	150.736	2.416	3.731	115.19	S
128	0.74	149.523	2.656	3.344	101.46	
129	0.74	149.866	2.492	4.466	170.82	S

ID	z	R.A.	Decl.	S/N	richness	previous detections
130	0.74	149.986	2.563	3.305	99.40	SO
131	0.74	150.024	2.688	3.388	103.83	S
132	0.74	150.051	2.302	3.821	128.61	S
133	0.74	150.116	2.705	4.030	141.57	SO
134	0.74	150.171	1.703	3.449	107.13	
135	0.74	150.296	2.378	3.196	93.72	S
136	0.74	150.547	2.803	3.212	94.55	S
137	0.76	149.559	1.647	3.054	88.26	
138	0.76	149.732	2.761	3.248	98.66	
139	0.76	149.839	1.684	3.468	111.29	
140	0.76	150.114	2.255	3.080	89.62	S
141	0.80	149.706	2.265	3.251	91.47	
142	0.80	150.040	2.652	3.906	126.31	
143	0.80	150.368	2.005	3.546	106.42	X
144	0.80	150.506	2.222	3.229	90.37	
145	0.80	150.538	2.148	3.097	84.12	O
146	0.80	150.580	2.652	4.594	169.49	S
147	0.80	150.702	2.769	3.850	123.09	

Table 4.2: As Table 4.1, but for clusters detected with the optical filter without lensing counterpart.

sponds to the latter, while at lower redshifts that angular position is outside the scanned area, because the filter would not be completely contained in the field of view (see Section 4.4.3). However, if we force our algorithm to analyse that region, we find an overdensity at $z = 0.24$ with coordinates (R.A. = 150.329, Dec. = 1.609), confirming the alignment of two different structures.

4.4.6 Characterizing our sample

In Fig. 4.15 we plot the number of detections obtained with our analysis (both optical and optical plus lensing, dashed and solid lines respectively), as a function of their optical S/N ratio together with the corresponding amount of detections already reported in literature. As expected, the percentage of 'matched' detections increases as a function of S/N. The only detection with S/N > 5 without a counterpart is actually near to a large structure observed by Scoville et al. (2007b) with a redshift mismatch of $\Delta z = 0.06$, only slightly larger than the limit we adopted.

To investigate the selection function of our method with respect to the mass, we plot in Fig. 4.16 the number of detections with respect to the X-ray masses estimated by Finoguenov et al. (2007). We include in this analysis only those X-ray clusters which lie inside the volume of our search, avoiding borders of the field and too high redshifts, as already discussed in Section 4.4.3. We see that the rate of X-ray clusters that are optically confirmed by our analysis increases with their mass, and it is above 50% for clusters of mass above $1.5 \times 10^{13} M_{\odot}$.

Finally we compare in Fig. 4.17 the corrected richness, corresponding to a measurement of the mass in galaxies, with the X-ray mass estimates made by Finoguenov et al. (2007). The proportionality between the two 'mass estimates' is clear although the scatter is very large. Note that the corrected richness is only a relative measure of the cluster stellar mass, in the sense that we do not quantify the physical mass scale. Note also that our filter has a fixed physical size, while the X-ray analysis was done inside an estimated r_{500} for each cluster, partially explaining the differences. The noise in the richness estimate for low-mass clusters is likely to be due to the field galaxies we observe inside our fixed spatial filter. This is expected given the error bars that show our analytic estimate.

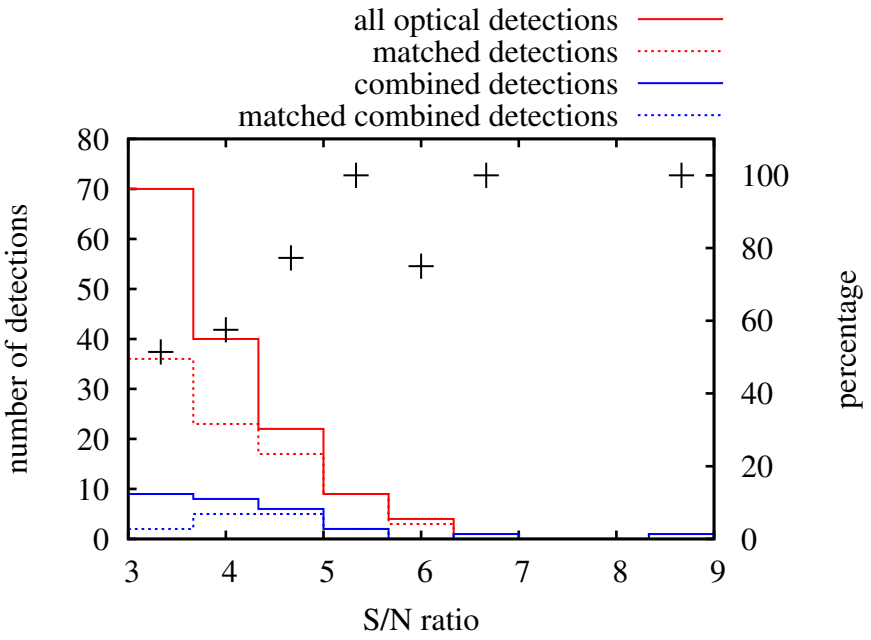

Figure 4.15: S/N distribution of all the significant optical detections (solid red), and of the detections with a counterpart in other analyses (dotted red). In blue we show the same distributions, but considering only combined (optical+lensing) detections. The black crosses indicate the percentage of our optical detections having at least one counterpart (scale on the right).

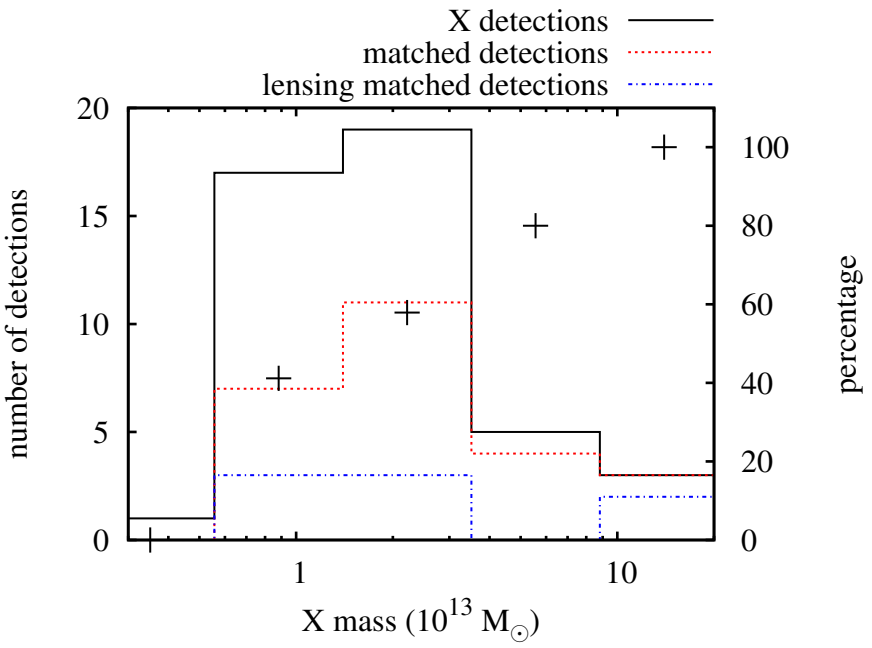

Figure 4.16: Mass distribution of all the X-ray detected clusters (in black), of the clusters we detect in the optical analysis (in red), and of those we detect in both optical and lensing analyses (in blue). The black crosses indicates the percentage of clusters detected in our optical analysis.

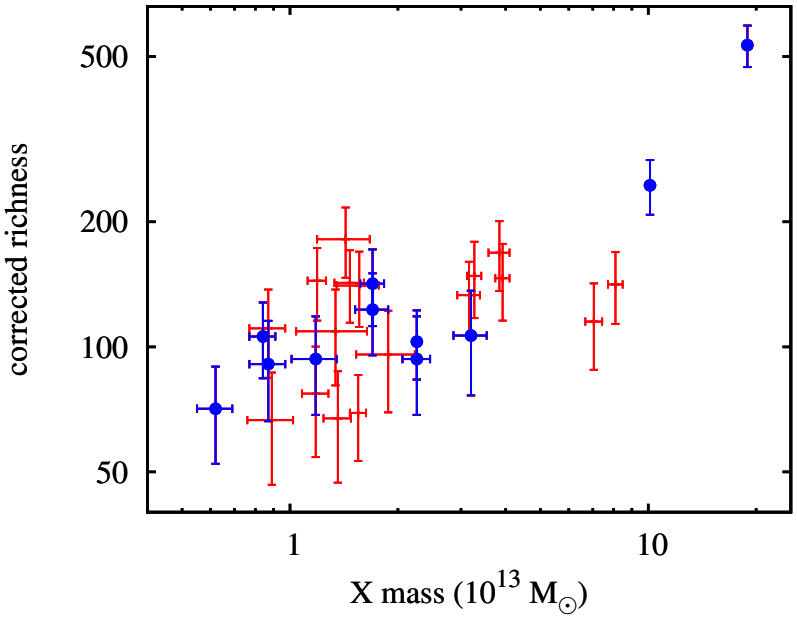

Figure 4.17: Correlation between the X-ray mass derived by Finoguenov et al. (2007) and our corrected richness. In red we show all the optical detections, in blue the lensing-confirmed clusters.

4.4.7 Summary and conclusions

In this Chapter we have presented an optimal linear filtering technique for the detection of galaxy clusters from optical and weak lensing data. The filter was first presented by Maturi et al. (2005) for weak lensing analysis and we now use the same formalism to extend it to optical data. The filter relies on a physical model for clusters and thus it accounts for all the known properties of galaxy clusters which can be inferred from photometric observations. Different information can be included and in this work we decided to restrict our analysis to the following observed properties of the galaxies: positions, magnitudes in one band, and photometric redshifts if present. The algorithm starts from a galaxy catalogue with the mentioned physical quantities and creates a map of values for the richness parameter Λ at different redshifts. The peaks of these maps represent the possible locations of galaxy clusters. The redshift of the clusters is determined by maximizing the likelihood of the data as a function of the redshift of the cluster model.

Our algorithm offers a number of improvements over existing filtering methods. Photometric redshifts are included in a flexible way, that adapts itself to the precision of the measurement of each galaxy, avoiding any sharp cut of the catalogue in redshift slices. It uses results from observed galaxy clusters to build a suitable model for cluster galaxy distributions. The richness calculated by the algorithm at the position of the cluster is a measurement of the number of galaxies in the detected cluster, that can be corrected for redshift dimming. Moreover, from the expected spatial profile and luminosity distribution of the clusters, and from the observed field population, it is possible to evaluate the noise of the detection and its significance. We tested the algorithm with numerical simulations, probing that it is able to obtain an unbiased estimate of the richness of the cluster, with an uncertainty predicted analytically from the model, and also its redshift.

We applied both the weak lensing and the optical filters to real data, obtaining a catalogue of candidate galaxy clusters for the COSMOS field. We presented a catalogue of 27 lensing-confirmed clusters, 11 of which do not have any previous detection in the literature. For the subsample of objects that have been analysed through their X-ray emission, we found that we are able to detect more than 50% of the clusters with an X-ray mass over $1.5 \times 10^{13}\ M_{\odot}$ and we found a good correlation between our galaxy richness

parameter and the cluster mass determined from X-ray temperature.

Chapter 5

Evaluating the performances of the Euclid mission

5.1 Galaxy density and photometric redshift precision

5.1.1 Introduction

In this Section we apply the SkyLens code, described in Sect. 3.2, using the expected characteristics of the Euclid mission, which we introduced in Sect. 2.4. As we already said in Sect. 2.3, some of the most severe requirements to perform precision cosmic shear measurements concern the galaxy angular density and the photometric redshift precision. We simulate realistic observations of patches of the sky, and submit them to codes normally used to analyse real observations, to verify that the instrumental and survey setups are suitable for the scientific purposes of the mission. The reference article for this Section is Bellagamba et al. (2012).

The telescope and detector characteristics of EUCLID are folded into SkyLens. The key parameters for the simulations are listed in Tab. 5.1. Taking into account all reflections and transmissions in the telescope design (five mirrors and a dichroic) and the detector/filter QEs, the total throughput varies as a function of wavelength as shown in Fig. 5.1. For producing the simulated images, we use system PSF models that are constructed on the basis of the expected optical PSFs and Attitude and Orbit Control System

parameter	value
Diameter	1.2 m
Obscuration	11 %
Pixel scale	0.101" (VIS)
	0.3" (NIP)
R.O.N.	$5e^-$ (VIS)
	$7e^-$ (NIP)
dark current	2 e^-/pixel/hr (VIS)
	0.05 e^-/pixel/s (NIP)
Throughput	see Fig. 5.1

Table 5.1: Key parameters used in the simulations.

pattern (AOCS), taking into account also possible additional detector effects. As an example, we show a system PSF model at 800nm in Fig. 5.2.

5.1.2 Galaxy angular density

To estimate correctly the number counts in the Euclid exposures, it is important to describe in a realistic way the contribution of the background light to the photon counts. The principal background contributor in the wavelength range where EUCLID will observe is the zodiacal light. This is the light of the sun reflected from the dust in the inner solar system. A model for the zodiacal light near the North-Ecliptic-Pole (NEP) has been derived by Leinert et al. (1998) and later revised by Aldering et al. (2004). The spectrum can be represented by a 'broken log-linear relation', given by

$$log_{10}[s_\lambda(\lambda)] = \begin{cases} -17 - 755 & (400 \leq \lambda \leq 610) \\ -17 - 755 - 0.730(\lambda - 610) & (610 \leq \lambda \leq 2200) \end{cases} , \quad (5.1)$$

where λ is given in nanometers and $s_\lambda(\lambda)$ is in erg cm^{-2} s^{-1} Å^{-1} . This spectrum corresponds to a 'low' background level and is normalized to a V-band surface brightness of 23.3 mag arcsec^{-2}. We define an 'average' background level as that normalized to a V-band surface brightness of 22.7 mag arcsec^{-2}. We also define the 'high' background level, which is normalized to a V-band surface brightness of 22.1 mag arcsec^{-2} . The 'average'

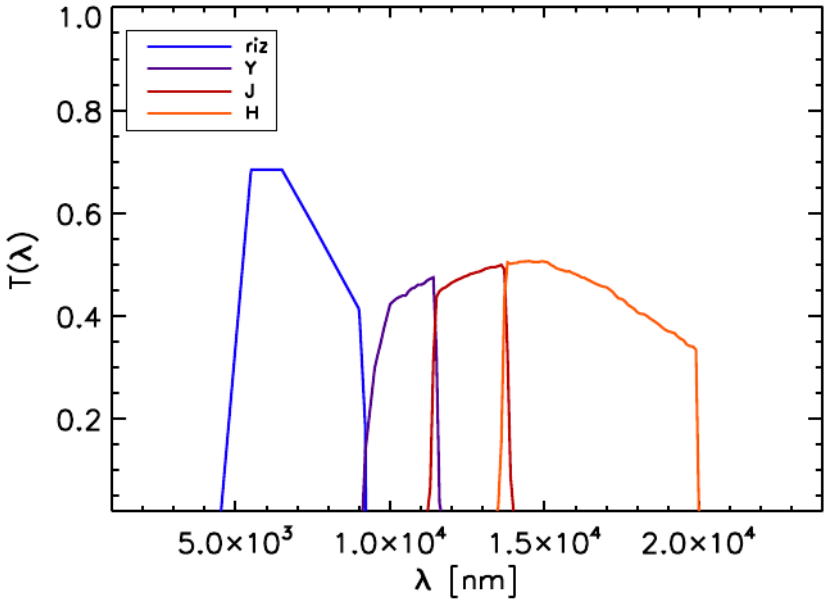

Figure 5.1: Total throughputs of the EUCLID visual (blue line) and NIP (magenta, red, and orange lines) channels.

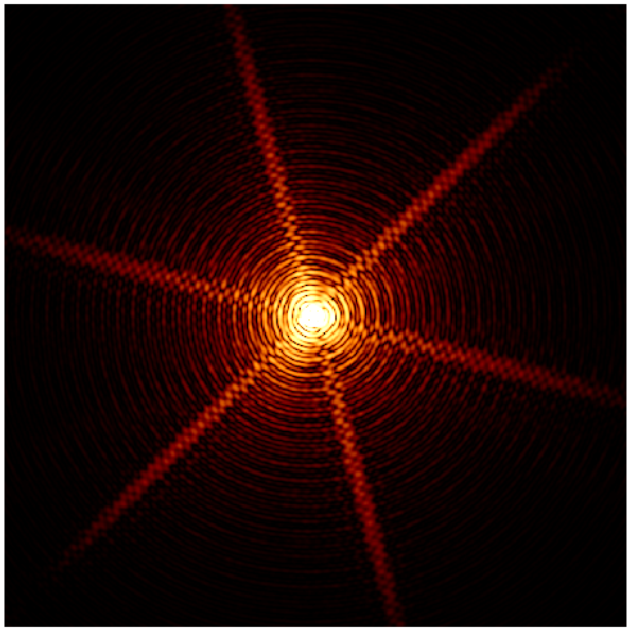

Figure 5.2: A model PSF used in the image simulations (λ = 800nm).

background is therefore 1.74 times brighter than the 'low' background level, while the 'high' background level is 3.03 times brighter than the 'low' background. In terms of AB magnitudes, the above spectra provide a surface brightnesses in the wavelength range covered by the VIS broad band (riz) of 22.95, 22.35, and 21.75 for the 'low', 'average', and 'high' background cases, respectively. Similarly, estimates of the brightness of the sky are derived also for the infrared bands.

Some examples of our simulations in the VIS band are shown in Fig. 5.3. The upper panels show a patch of the sky, which corresponds to a region of 85" × 45", as it might be seen by EUCLID combining two (left) and four (right) exposures of 540s each, assuming the 'average' sky background. Instead, the bottom panels show the same patch observed combining four exposures, assuming the 'high' (left) and the 'low' sky background levels, respectively. Similar images have been analyzed to assess the number counts of galaxies useful for the weak lensing analysis. These are defined as galaxies with signal-to-noise ratio (SNR) larger than 10 and size larger than 1.25 times the PSF size. The SNR and the size of each galaxies were measured in the EUCLID simulated images using the software SExtractor (Bertin & Arnouts, 1996). The resulting number counts for a range of effective exposure times and possible sky background levels are shown in Fig. 5.4.

5.1.3 Multi-band photometry

In the rest of this Section, we will assess what is the achievable precision in the redshift estimation using Euclid space data in connection with multi-band optical ground-based observations. This kind of study has been performed up to now by creating multi-band mock galaxy catalogues with MonteCarlo techniques, considering the nominal depth of the survey, and thus the expected error on the photometric measurements (Abdalla et al., 2008; Bordoloi et al., 2010). With this work we study in deeper detail this process, by using complete simulations, from the images of the galaxies in different observing bands to the photometric redshift estimation. In this way, the photometry of each galaxy has a more realistic uncertainty and we can take into account more sources of error in the measurements, such as those due to proximity of the galaxies. We start from a 800" × 800" image of a simulated sky observed with the Euclid experiment (see Fig. 5.5). We underline that this broad band

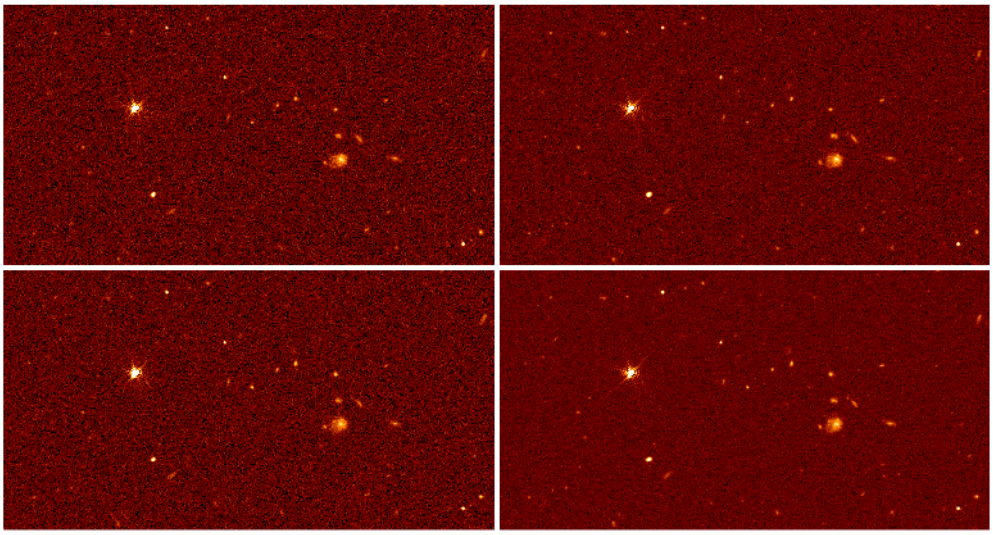

Figure 5.3: Simulated EUCLID-VIS observations of a patch of the sky (8" × 45"). The upper panels refer to different exposure times: the left and the right panels show the result of combining two and four 540s-long exposures, respectively. The left and the right bottom panels show the same field assuming a high- and a low-level sky background.

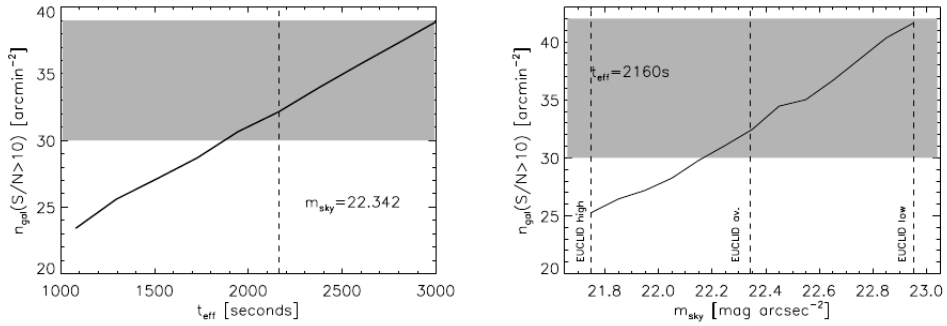

Figure 5.4: Number counts of galaxies useful for weak lensing in future EUCLID observations. These are galaxies with SNR>10 and size larger than 1.25 times the PSF size. In the left panel, the counts are shown as function of the effective exposure time for the 'average' sky background level. In the right panel, we fix the exposure time to 2160s and vary the sky brightness between 22.95 (low background) and 21.75 mag/arcmin^{-2} (high background). The goal is to design the survey strategy such that the counts are in the grey-shaded region to match the EUCLID requirements, 30-40 objects per arcmin^2.

Figure 5.5: Simulation of an observation with Euclid in the RIZ band. The picture has a side of 400 arcsec. The exposure time is 1800 sec.

is proposed by the Euclid experiment to maximize the galaxy density used in cosmic shear measurements, but it is practically useless in photometric redshift determination, so the results of this work do not depend strongly on this choice. We then simulate the near-infrared coverage of the same sky, in particular considering the three Euclid NIP bands (Y J H). Finally, we simulate the observation of the same sky with different ground-based instruments and different observing conditions. In the following, we explain how we analyse these images to extract the redshift estimation. In Fig. 5.6 we show as an example the appearance of the same galaxy in different observing bands and conditions.

Our intent is to submit the images to software pipelines usually adopted for the analysis of real observations, to thoroughly simulate the extraction

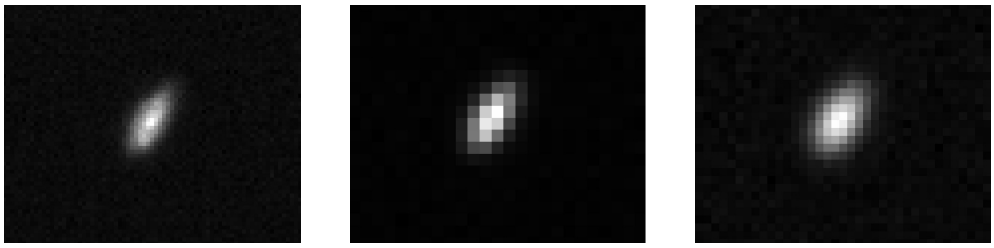

Figure 5.6: A galaxy observed with Euclid in the RIZ band (left), with Euclid in the H band (centre), and with Subaru in the i band (right) .

of a redshift estimate from images taken with different instruments. A fundamental step to get a good photometric redshift estimate is to correctly measure the colours from multi-band observations of the same galaxy. As the images have different PSFs and different levels of noise, measuring naively the magnitude in each band in a matched aperture is not appropriate. Instead, we use the Colorpro software, presented by Coe et al. (2006), that measures all the magnitudes in the same aperture, and then corrects them for the different PSFs. This software has been used in many experiments, including some that combine space-based and ground-based data, as in our case (e.g. Umetsu et al., 2010; Rafelski et al., 2009).

In particular, we use the image of the galaxy in the RIZ band as the detection image, where the galaxy is detected and its aperture defined. To catch possible optical dropouts, we detect galaxies also in the H-band image, and we merge the two catalogues. Then the images in all the bands are remapped to the higher resolution of the RIZ image, and SExtractor MAG_ISO magnitudes are measured from these images, considering the apertures previously defined. We underline that all the apertures we use for the photometry are drawn from the space-based RIZ image, to take advantage of the better PSF. Then, to estimate the amount of flux lost in each band due to the worse seeing, the RIZ image is degraded to the PSF of each band, and its new MAG_ISO measured. The difference between this magnitude and the MAG_AUTO measured on the original RIZ frame is the desired correction. All magnitudes stated in the following are AB magnitudes.

An example of the precision in the colour measurements obtained applying Colorpro to our simulations is shown in Fig. 5.7, where we plot the mean offset of measured colours with respect to the input values. We con-

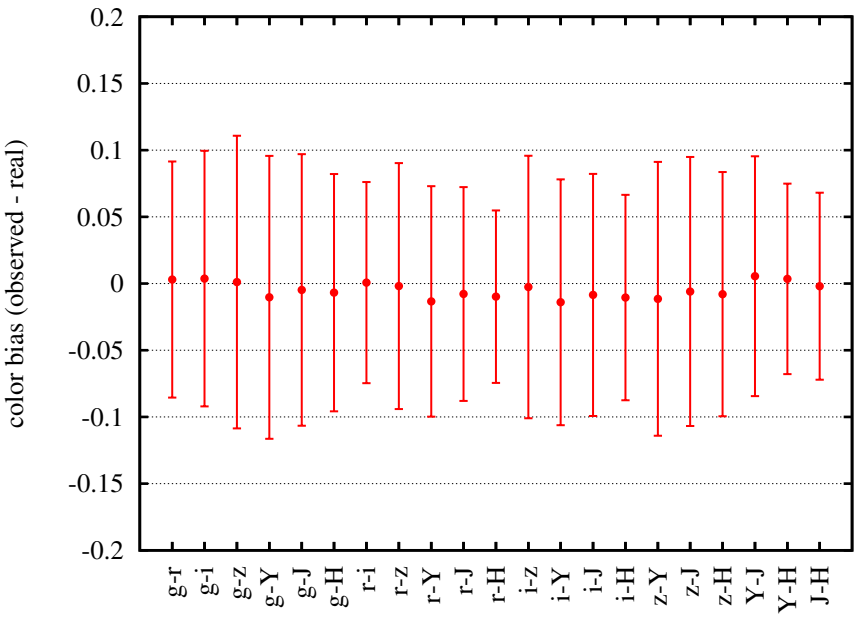

Figure 5.7: Precision in the colour measurement of our simulated images (Euclid + LSST) by Colorpro. For every couple of bands we show the mean of the colour bias over the galaxy sample, while the error bars represent the standard deviation of the measurement. This test has been performed observing one galaxy at a time, to avoid pollution from neighboring galaxies.

sidered combinations between the four space-based bands (RIZ Y J H) and the ground-based bands that will be used in the photometric redshift estimation (*griz*). For the purpose of this plot, we simulated the observation of a small field around one galaxy at a time, to avoid blending and pollution and thus check the robustness of the photometry estimation process in an ideal case. We see that in this conditions the mean bias is negligible and the standard deviation is below 0.1 magnitudes. We then focus on a single colour (*g*-J) and plot in Fig. 5.8 the error in the colour measurement as a function of the photometric uncertainty for each object in the sample. We see that most of the objects have very small values of bias and that the statistical standard deviation is in good accordance with the uncertainty quoted by the program. This means that the implemented method is able to compensate for the different PSFs and pixel sizes of the images.

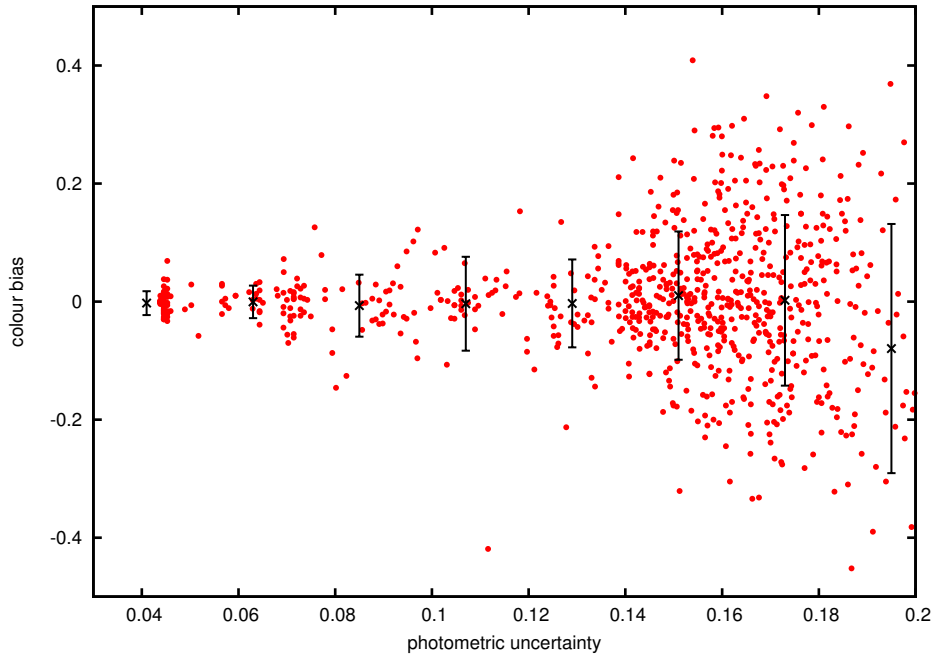

Figure 5.8: Precision in the colour measurement (LSST r band - Euclid J band) of our simulated images by Colorpro. For each object we show the colour bias as a function of the photometric uncertainty. The red bars show the mean and the standard deviation of the measurement in bins of photometric uncertainty.

5.1.4 Redshift estimation

With the procedure described above, we have obtained realistic multi-band catalogues, with luminosity measurements in the 4 space-based bands (RIZ,Y,J,H) and in the proper ground-based optical bands. Now we use these catalogues in the proper estimation of photometric redshifts for the galaxies. As before, we perform this task applying a method that is developed for and usually applied to real data. Different techniques have been proposed to extract the redshift estimate of a galaxy from its multi-band photometry, such as template fitting (e.g. Bolzonella et al., 2000; Benitez, 2000; Feldmann et al., 2006) and training-set methods (e.g. Collister & Lahav, 2004; Way & Srivastava, 2006). They are probed to obtain similar results (Abdalla et al., 2011), with the former being more flexible, because they do not need to be trained with a complete spectroscopic subsample. For this work we use the BPZ software (Benitez, 2000), the implementation of a Bayesian template-fitting method, that has been applied to real data in many works (e.g. Mei et al., 2006; Menanteau et al., 2006). It requires as input a set of galaxy template SEDs, from which the program calculates the colours when observed through the considered filters at different redshifts. The probability of a galaxy with colours C of being at redshift z is then evaluated applying

$$p(z|C, m_0) \propto \sum_T p(z, T|m_0) p(C|z, T) \tag{5.2}$$

where m_0 is the magnitude used to compute the prior and T is the index that runs over the templates. In particular, $p(z, T|m_0)$ is the prior given by the i band distribution, as observed in the Hubble Deep Field North, while $p(C|z, T)$ is the likelihood of the colours given the template and the redshift, calculated via a χ^2 approach,

$$-\log p(C|z, T) \propto \chi^2(z, T, a) = \sum_\alpha \frac{(f_\alpha - a f_{T\alpha})^2}{\sigma_{f_\alpha}^2}, \tag{5.3}$$

where $f_{T\alpha}$, f_α and σ_{f_α} are the theoretical flux of the template, the observed flux and the noise in the band α, respectively, and a is the normalisation of the template. The BPZ software has been proven to perform competitively with other methods in Abdalla et al. (2011). Further details and a little modification of the original algorithm can be found in the next Subsection.

In our case, we know the SEDs of our simulated galaxies, so we give the 'right' template SEDs to the program. This set, which we already introduced in Sect. 3.2, is formed by the six templates used by Benitez (2000), plus the two added by Coe et al. (2006) to fit the bluest galaxies of the HUDF sample. This is an ideal case, as in photo-z template-fitting algorithms the correct choice of the template set is one of the critical steps (Yee, 1998; Rowan-Robinson et al., 2008). The input magnitude that BPZ uses to compute the prior is the one measured in the i band by the appropriate ground-based instrument.

5.1.5 Correction to BPZ algorithm

In Section 3.2 of Benitez (2000), the author introduces the normalisation parameter a of the template SED as a nuisance parameter in the Bayesian formalism. The probability of observing the colours C given the template T and the redshift z is obtained by integrating the χ^2 likelihood (see Eq . 5.3) over a, instead of considering only the best-fit value a_{min} as in the simple maximum-likelihood approach. Thus, we have

$$p(C|z,T) \propto \int da\, p(a|m_0)\, p(C|z,T,a), \tag{5.4}$$

where m_0 is the magnitude value used as a prior. The prior of a with respect of m_0, $p(a|m_0)$, is conservatively assumed flat. Evaluating the integral, he gets

$$p(C|z,T) \propto F_{TT}^{-1/2}\, \exp\left[-\frac{\chi^2(z,T,a_{min})}{2}\right], \tag{5.5}$$

where

$$F_{TT} = \sum_{\alpha} \frac{f_{T\alpha}^2}{\sigma_{f_\alpha}^2}. \tag{5.6}$$

Eq. 5.5 and 5.6 mean that the value of $p(C|z,T)$ depends over the ratios of the expected fluxes over the observed errors in the different bands and thus over the input normalisation of the templates (i.d., their value with $a = 1$). Without any indication on the way one should normalise $f_{T\alpha}$, this introduces an arbitrary factor in the calculation.

To overcome this problem, one can repeat the calculations assuming this time that the template is re-normalised at its best-fit parameter. We define

our new templates $f'_{T\alpha} \equiv a_{min} f_T$ and we introduce the nuisance parameter $a' \equiv a/a_{min}$. In this way, we avoid any arbitrariness in the calculation, by defining the templates and the nuisance parameter in a physically motivated way. With the new definitions, Eq. 5.5 becomes

$$p(C|z,T) \propto F'^{-1/2}_{TT} \exp\left[-\frac{\chi^2(z,T,a'_{min}=1)}{2}\right], \tag{5.7}$$

where F'_{TT} is the analogue of F_{TT} with the new template, and the value of a' that maximizes the likelihood is unity by definition. We can now re-express Eq. 5.7 in terms of the original templates. We have

$$\sqrt{F'_{TT}} = \left(\sum_\alpha \frac{f'^2_{T\alpha}}{\sigma^2_{f_\alpha}}\right)^{1/2} = \left(\sum_\alpha \frac{a^2_{min} f^2_{T\alpha}}{\sigma^2_{f_\alpha}}\right)^{1/2}. \tag{5.8}$$

We now make use of Eq. 8 of Benitez (2000)

$$a_{min} = \frac{F_{OT}}{F_{TT}}, \tag{5.9}$$

where

$$F_{OT} = \sum_\alpha \frac{f_{T\alpha} f_\alpha}{\sigma^2_{f_\alpha}}. \tag{5.10}$$

Inserting Eqs. 5.9 and 5.8 in Eq. 5.7, and using Eq. 5.6, we get

$$p(C|z,T) \propto \frac{\sqrt{F_{TT}}}{F_{OT}} \exp\left[-\frac{\chi^2(z,T,a_{min})}{2}\right]. \tag{5.11}$$

Now, in the term $\sqrt{F_{TT}}/F_{OT}$ the input normalisation of the original template cancels out, as we desire. Moreover, we can rewrite (see Eq. 7 of Benitez (2000))

$$\chi^2(z,T,a_{min}) = \sum_\alpha \frac{f^2_\alpha}{\sigma^2_{f_\alpha}} - \frac{F^2_{OT}}{F_{TT}}. \tag{5.12}$$

By comparing Eq. 5.11 and Eq. 5.12 we see that, given the data, the errors and the χ^2 goodness of fit with an unknown template, we have a certain determination of the probability integrated over the normalisation parameter of the template.

In practice, we have modified the BPZ program by inserting Eq. 5.11 in the place of Eq. 5.5. This change eliminated some issues we had especially with late-type galaxies at high redshift, reducing by more than a

factor of two the number of catastrophic failures. We note here that Coe et al. (2006) preferred to eliminate the $F_{TT}^{-1/2}$ term in 5.5, reporting numerical instabilities. Actually this would give results very similar to ours, because the $\sqrt{F_{TT}}/F_{OT}$ term in Eq. 5.11 is almost irrelevant with comparison to the maximum-likelihood exponential.

5.1.6 Catalogue cleaning

Photometric redshift estimation suffers from colours degeneracy for galaxies at distinct redshifts. Although the Bayesian prior implemented in Eq. 5.2 reduces this problem, e.g. avoiding that very brilliant galaxies are given high redshift estimates, it does not solve it completely. When two templates at different redshifts have very similar theoretical colours, photon noise will sometimes move the observed magnitudes of a galaxy towards the wrong template, and the redshift extracted from the algorithm will be catastrophically wrong. Because of this, to match the strong requirements on the precision in the photometric redshifts for cosmic shear surveys, we need to clean our catalogue from galaxies that are likely to have a wrong redshift estimate. A way to do so is looking at the full redshift probability distribution of the galaxy, instead of considering only the most probable value. Galaxies that have colours in regions of degeneracy, i.e. that lie near the theoretical colours of templates at different redshifts, will have a multi-peaked probability distribution. Galaxies that have a controversial photometry, for example because are polluted by brilliant neighbours, will not show a single strong peak, but a diffuse distribution with much of the probability outside the vicinity of the most probable redshift. Thus, measuring the integral of the redshift probability distribution of the galaxy in the proximity of the most prominent peak will give an indication of how certain is the redshift estimation. This can be done via the ODDS parameter calculated by BPZ. It is defined as

$$ODDS = \int_{z'-2\Delta_z}^{z'+2\Delta_z} p(z), \tag{5.13}$$

where z' is the most probable redshift value, and Δ_z is a free parameter, set by default in BPZ to $0.067(1+z')$. ODDS is thus a measure of how much the distribution is concentrated around the principal peak. This is indeed a powerful way to assess the goodness of the estimate, as can be seen in Fig. 5.9,

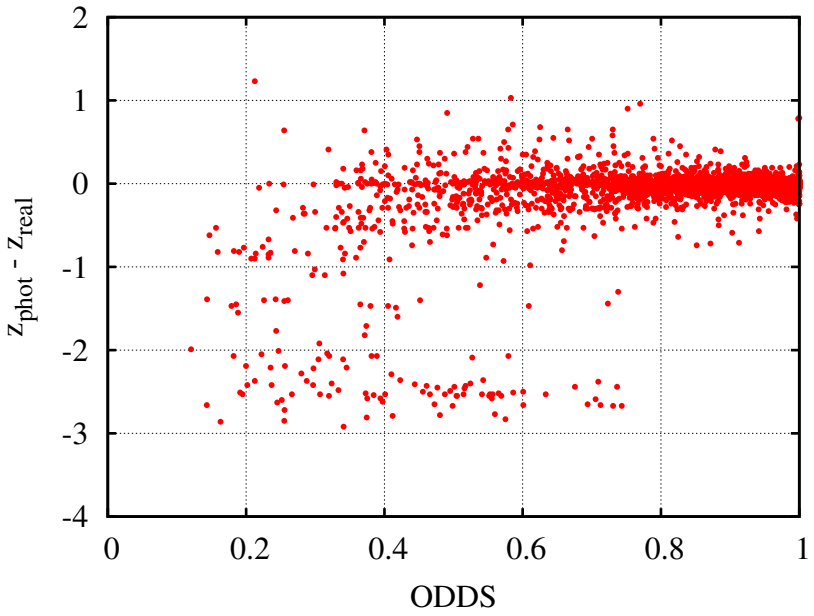

Figure 5.9: Example of the power of the ODDS parameter in the identification of catastrophic failures. The error of the photometric redshift estimate is plotted against the ODDS value computed by BPZ. Data taken from the 100 sec observation with the ideal telescope, u band excluded (See Section 3.)

where we plot for each galaxy the error in the redshift estimate against the ODDS parameter calculated from its $p(z)$. The fraction of galaxies with a catastrophic redshift estimate decreases rapidly as ODDS increases.

If ODDS > 0.95, it means that the distribution is more peaked than a Gaussian with a width equal to Δ_z. This may seem a reasonable cut for the catalogue, and has been used in some works (Coe et al., 2006; Hildebrandt et al., 2008), while Erben et al. (2009) preferred a looser limit of 0.90. In the example of Fig. 5.9, we see that catastrophic failures become rarer for galaxies with an increasing ODDS value, and disappear for galaxies with ODDS > 0.80. In general, we need to trade off the amount of galaxies to be kept in the catalogue with the mean accuracy of the redshift estimation. In the following, we will try to optimise the selection, by looking at how the standard deviation of the redshift measurements varies as a function of the

ODDS value used as a minimum threshold. An optimised selection is crucial in cosmic shear observations, as it is important to consider as more galaxies as possible to reduce the noise induced by the intrinsic shape of the galaxies.

In particular we will measure the precision of the redshift estimates calculating the normalised standard deviation

$$\langle \frac{\sigma_z}{1+z} \rangle = \frac{1}{N} \left[\sum_{i=1}^{N} \frac{(z_{i,real} - z_{i,phot})^2}{1 + z_{i,phot}} \right]^{1/2}, \tag{5.14}$$

where the index i runs over the N galaxies of the sample in the redshift range $0.3 < z < 3$. We fix the requirement on the precision in the redshift measurement to $\langle \frac{\sigma_z}{1+z} \rangle < 0.05$, that is considered valid to extract cosmological information from cosmic shear experiments, such as Euclid (Laureijs et al., 2009). We will measure how many galaxies we must eliminate from the sample via a selection based on the ODDS parameter to fulfil this condition. As an example, in Fig. 5.10 we plot the value of $\langle \frac{\sigma_z}{1+z} \rangle$ as a function of the ODDS value used as a lower threshold, for one of the simulated observations. In practice, we continue to throw away galaxies from the sample according to their ODDS value until the normalised standard deviation of photo-z estimates goes below 0.05. In this procedure, we will unavoidably eliminate some of the galaxies with a good photo-z estimate, and we will as well keep in the catalogue some outliers, as the indication given by ODDS is obviously only valid in a statistical sense. Nonetheless, the mean standard deviation drops almost monotonically increasing the ODDS threshold, so in this way we are in the end able to extract a robust galaxy sample from the data.

For every survey setup our main result will be the ratio of the number of galaxies that are kept in the catalogue after the cut over the total number of galaxies in the initial sample. We will take into account only galaxies with a measured RIZ magnitude below the declared limit of the Euclid mission (24.5). Although with a Euclid-like setup we are able to go slightly deeper than this limit in the optical imaging observation, some tests we did including dimmer galaxies proved not to improve our results, probably because they suffer from poor infrared photometry, so we decided to concetrate on a RIZ-magnitude limited sample. To increase the statistic, we create 1000 random realisations of each catalog via bootstrap resampling, and report the mean of the measurements.

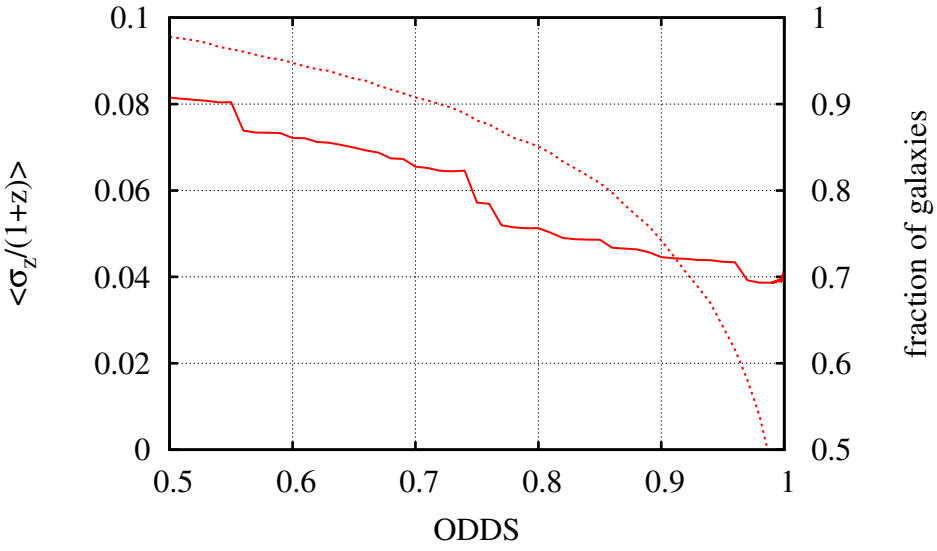

Figure 5.10: Normalised standard deviation (solid line) and fraction of available galaxies (dashed line) as a function of the ODDS value used as the minimum threshold for the photo-z catalogue. Data taken from the 100 sec observation with the ideal telescope, u band included (See below.) In this case, we must cut the catalogue at ODDS $\simeq 0.80$ and eliminate $\simeq 15$ % of the sample to get to the required precision, $\langle \frac{\sigma_z}{1+z} \rangle = 0.05$.

	ideal	CTIO	PAN-STARRS	LSST
Area ratio	1.	0.149	0.193	0.628
u eff.	1.	-	-	0.533
g eff.	1.	0.539	0.478	0.713
r eff.	1.	0.615	0.679	0.786
i eff.	1.	0.669	0.706	0.813
z eff.	1.	0.484	0.316	0.360
Pixel size	0.2	0.27	0.3	0.2

Table 5.2: Comparison of the performances of the considered telescopes and optics. Efficiencies of the ideal telescope are equal to 1 by definition.

5.1.7 Dependence on survey parameters

In order to obtain general results that do not depend on the specific instrument or survey, we first simulate some optical ground-based multi-band observations taken with a fictional ideal instrument. This instrument has a very large collecting area (diameter = 8.2 m), a pixel angular size of 0.2 arcsec and does not suffer from any losses due to mirrors and optics. The considered filters are the standard SDSS *ugriz*. The performances of this unrealistic setup are compared with the ones of real telescopes considered in this paper in Tab. 5.2. Given the exposure time in a certain band with the ideal telescope, it is easy to calculate what is the time needed to reach the same depth with another telescope, considering the efficiency and the area of the instruments.

We underline that all the results presented in the following are obtained using data from the 4 space-based bands (RIZ Y J H), plus the indicated ground-based data.

The quality of the photometry, and thus of the photometric redshifts, depends primarily on the depth of the survey. In this view, we simulate a set of observations with a constant exposure time in each *griz* band. (The importance of the u band will be studied separately in the next Section.) The results of the simulation are shown in Fig. 5.11, where we show the standard deviation of the total sample (i.e. formed by all the galaxies with RIZ magnitude < 24.5, without any selection based on ODDS), and the fraction of galaxies that are kept in the catalogue after a selection analogue to Fig. 5.10

is performed. We see that the standard deviation of the whole sample does not improve significantly as the exposure time increases, but there is a great enhancement of the number of available galaxies after the cut is performed. This happens because the standard deviation of the total sample depends mainly on catastrophic failures due to redshift-colours degeneracies. These are only partially avoidable improving the quality of the photometry, because they are an intrinsic drawback of the photometric redshift estimation. Above this, we have a number of galaxies with poor photometry due to bright companions or to blending with close dim neighbours. In this case, the observed flux is the sum of two different objects, and thus the observed SED will likely not correspond to any of the two physical states. Luckily, the selection via the ODDS parameter allows us to reject these galaxies. The improvement in the fraction of available galaxies after the selection for longer exposures is then given by the more precise photometry that we obtain. Reducing the errors on the measured magnitudes, the non-pathological galaxies (i.e. not exactly on a redshift-space degeneracy and without close companions) will be assigned a redshift with a smaller uncertainty. Thus they will have a more peaked redshift probability distribution and a bigger ODDS parameter. This will avoid them from being confused with catastrophic failures when the selection is performed. We note that with the deepest exposure, 1000 sec, just the 14% of the total sample of galaxies brighter than 24.5 in the RIZ band must be excluded to meet the requirements in the redshift precision.

It is sometimes possible to extend the observed wavelength range using the u band. This band is more time consuming then the others due to the increased absorption of the atmosphere with respect to the redder wavelengths and to the lower quantum efficiency of the CCD detectors. It is then usually necessary to trade off the availability of data at shorter wavelength with the loss of depth in the other bands. To test the importance of u-band data, we run the BPZ algorithm on the same images of the previous Section, but with an additional u-band observation of the same exposure time. The results are shown in Fig 5.12.

We see that even the shorter exposure that includes the u band (50 sec, corresponding to a 10-σ depth equal to 23.5) gives a total redshift standard deviation lower than with the deepest exposure with the $griz$ bands only. This result is not surprising as the power of the u band in photometric redshift

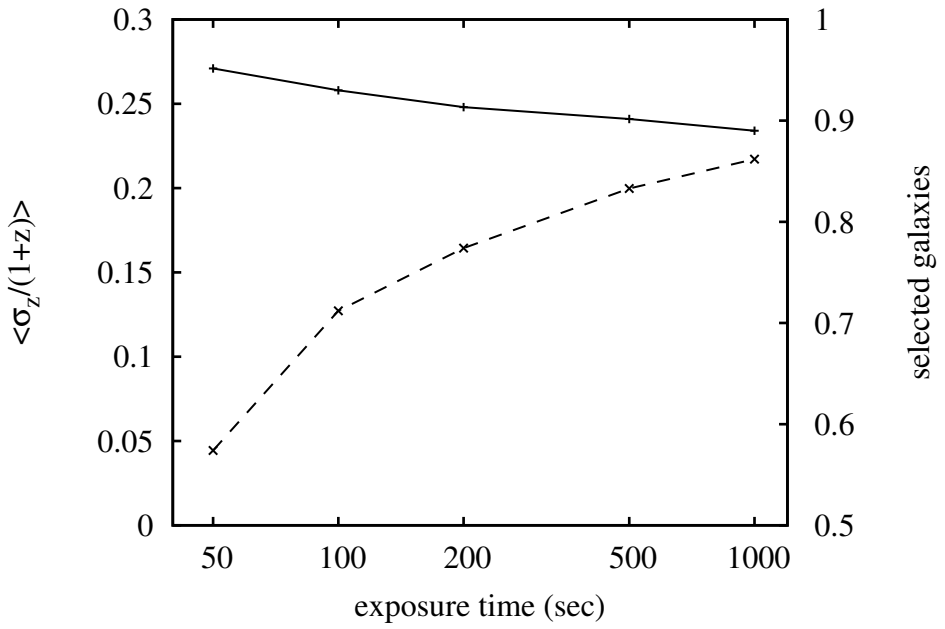

Figure 5.11: Quality of the photometric redshifts obtained with the *griz* bands of the ideal telescope as a function of the exposure time. The solid line represents $\langle \frac{\sigma_z}{1+z} \rangle$ as a function of the observing time, while the dashed line represents the fraction of galaxies with observed RIZ magnitude < 24.5 that are available after a cut is performed to reduce the redshift standard deviation below 0.05.

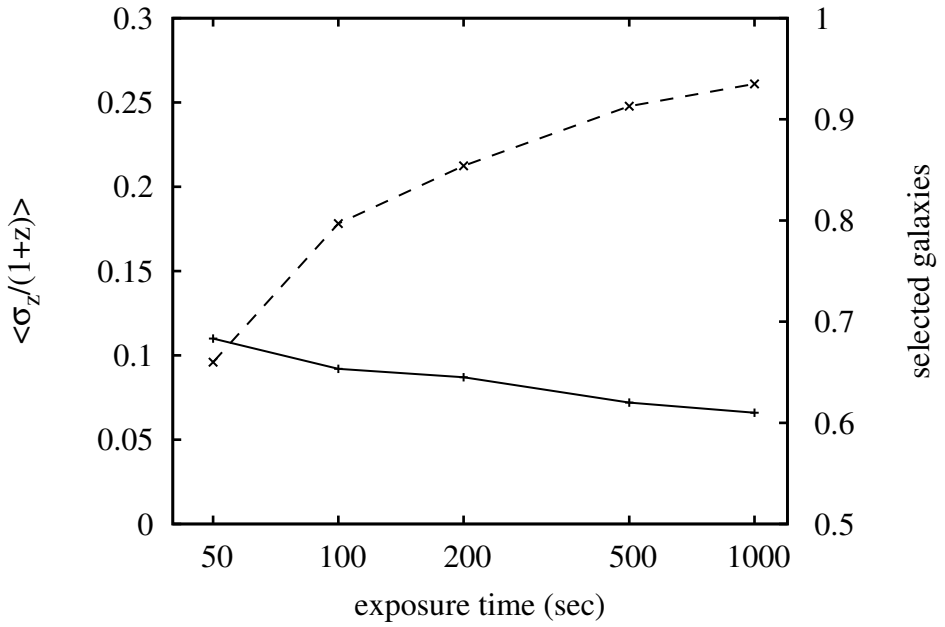

Figure 5.12: As Fig. 5.11, but including also the u band data.

estimation is widely recognised in the literature (Abdalla et al., 2008; Dahlen et al., 2010) . In particular, it helps in identifying galaxies at low redshift (z $\lesssim$ 0.5) thanks to the ultraviolet continuum, and the ones at z $\sim$ 3 when the Lyman break enters the blue side of the filter. The realisation of this can be seen in Fig. 5.13, where we plot the output redshift against the input one, both with and without using the u-band data. The reduction of catastrophic failures especially for galaxies at high redshift is very evident.

On the other hand, when considering the number of galaxies after the selection, we see that the results are closer to the ones with the *griz* bands only, thanks to the fact that many catastrophic failures are in that case pointed out and eliminated. The difference between the two setups is around 10 % of the total number of galaxies for the whole range of exposure times tested. With the longest exposures, the *ugriz* configuration permits to use around 94 % of the sample.

A problem that influences the quality of the photometry in ground-based observations is the blurring of the images that occurs as the light travels through the atmosphere. This is often quantified by the seeing, i.e. the

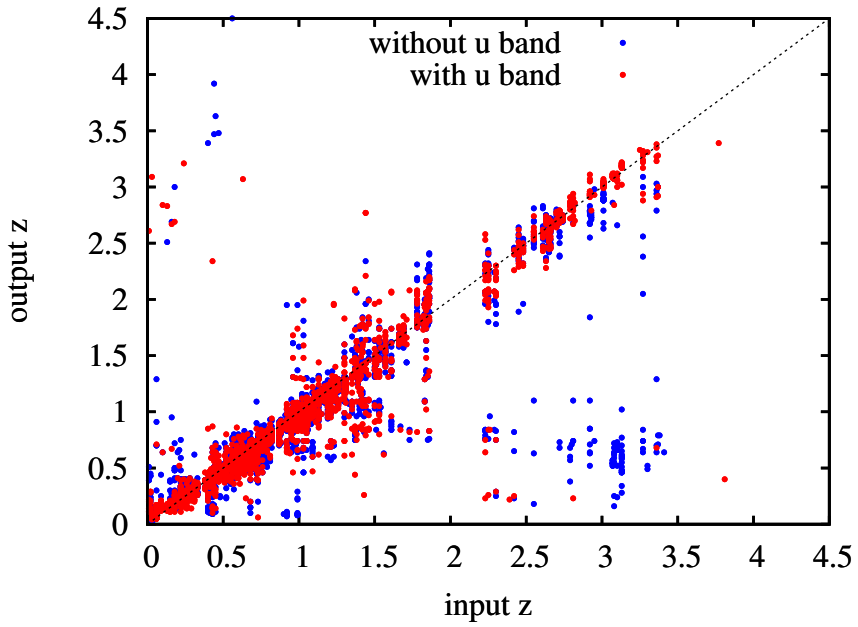

Figure 5.13: Estimated redshift against input redshift of the sample galaxies, considering the u band (red points) or discarding it (blue points). Data are taken from the observations with an exposure time equal to 100 sec per filter. The alignment of some group of points along the y-axis is caused by the repetition of the same galaxy in different sky positions, due to the limited amount of galaxies in the initial HUDF catalogue.

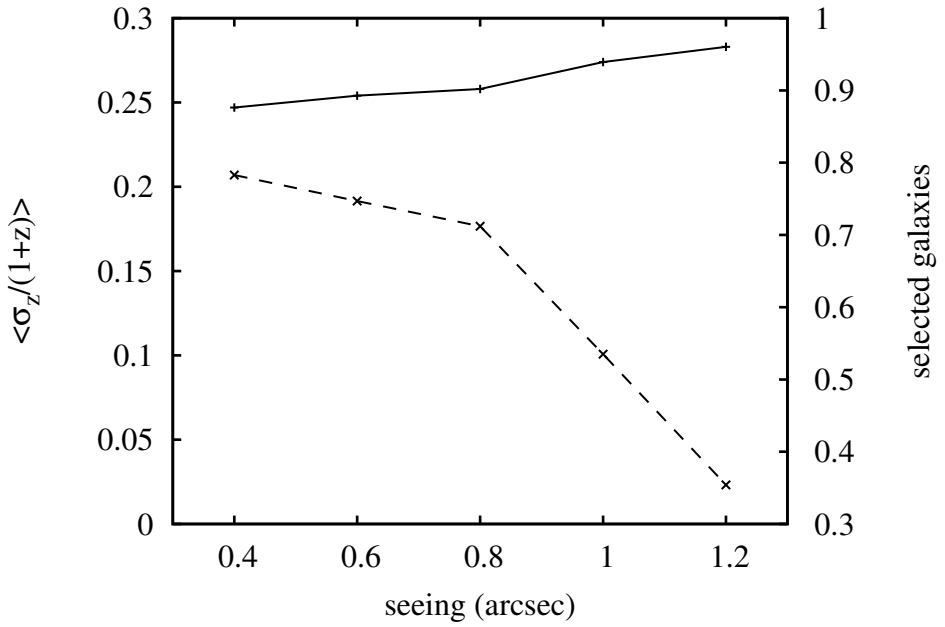

Figure 5.14: Quality of the photometric redshifts obtained with the *griz* bands of the ideal telescope as a function of seeing. The solid line represents $\langle \frac{\sigma_z}{1+z} \rangle$ as a function of the seeing, while the dashed line represents the fraction of galaxies with observed RIZ magnitude < 24.5 that are available after a cut is performed to reduce the redshift standard deviation below 0.05.

mean FWHM of the observed image of an intrinsic point-like source as a star. The method used to perform these simulations allows us to study the influence of this parameter on the precision of the photometric redshifts. In particular, we simulated five different observations of 100 sec per filter with the ideal telescope, with a seeing varying from 0.4 to 1.2. Results are shown in Fig. 5.14.

As in the case of the observing time, the change in the precision of the whole sample is not significant, while much more evident is the improvement in the number of available galaxies after the cut is performed. In fact, the number of galaxies with a certain redshift estimate can change by a factor of two depending on the quality of the atmosphere of the telescope site. This of course depends on the leakage of light between different galaxies that occurs if the seeing is greater than the angular distance between them. This creates

an observed SED without a corresponding galaxy template and thus with a poor redshift estimate.

In reality, many astronomical images are the sum of exposures taken in different moments and with different observing conditions, i.e. without a single seeing value that can be attributed to the whole image, as in our idealised work. Nonetheless, our results are a strong indication that good observing sites are required to perform the high-precision photometry needed for redshift estimation of deep and vast galaxy catalogues.

5.1.8 Planned surveys

Different projects are considered as possible ground-based counterparts of space-based weak lensing observations. We underline that any space-based mission that will observe both the northern and southern sky will need collaboration from different projects covering different regions of the sky. In some sense, some of the projects we study will not be competitors against each other in practice, because any space-based project will need a patchwork of collaborations around the globe.

In this work we consider three ground-based surveys: the Dark Energy Survey (DES), Pan-Starrs and the Large Synoptic Survey Telescope (LSST). The first two are officially supporting the Euclid mission, while the third is a case we chose in order to explore the potential of observations that consider the u band, following the results of the previous Section. In Table 5.3 we show the exposure times and the depths of the three surveys, calculated with our simulator.

- DES is a project that will begin in late 2011 with the Blanco 4-meter telescope at Cerro Tololo Inter-American Observatory, on which a new camera named Decam will be mounted at the scope (Abbott et al., 2005). It will observe 5000 square degrees in the South Galactic Gap for 525 nights in 5 years.

- Pan-Starrs is a new telescope, consisting of four mirrors, one of which is already working at Mount Haleakaka in Hawaii (Kaiser et al., 2002). It will observe the 30000 square degrees visible from the site. The exposure times of Table 5.3 are calculated from the ones declared for

Band	DES time	DES depth	PAN-ST. time	PAN-ST. depth	LSST time	LSST depth
u	-	-	-	-	1800	25.7
g	500	24.3	2400	25.6	2400	26.6
r	500	24.0	1200	24.9	5400	26.6
i	900	24.1	1200	24.5	5400	26.2
z	2100	24.0	1200	24.1	4920	25.7

Table 5.3: Exposure time and 10σ point-source sensitivities for the simulated surveys. Although some of the experiments foresee to observe also in the y band, we did not consider it because its wavelength range is well covered by the space Y band.

the ongoing 3-years 3π survey (Magnier et al., 2007), multiplied by $10/3$, to make them realistic for a 10-year project.

- LSST is a telescope that will start observing in 2017 from Cerro Pachon in Chile (Abell et al., 2009). It will observe the southern half of the sky. The observing times are the ones declared for the 10-year total observing project.

For each project, we perform the simulations keeping the seeing fixed at 0.8 arcsec. This is a reasonable value, that can be obtained in all three sites (Els et al., 2009; Schöeck et al., 2009; Tokovinin & Travouillon, 2006). The results are given in Fig. 5.15.

As expected from the simulations with the ideal telescope, there is a big difference between the two *griz* instruments and LSST, when concerning the standard deviation in the redshift measurement of the whole sample. On the other hand, the number of galaxies available after the selection differs a lot also between the two *griz* instruments. The longer exposure times of Pan-Starrs make its results almost the best achievable for a *griz* survey, with more than 80 % of the galaxies retained after the selection. The LSST-like survey is the only one that goes beyond 90 %, as could be guessed by the simulations with the ideal telescope. A detailed plot of the redshift standard deviation as a function of the fraction of galaxies kept in the sample can be seen in Fig. 5.16. From this figure it is possible to evaluate a possible trade-off between redshift precision and galaxy fraction for the different in-

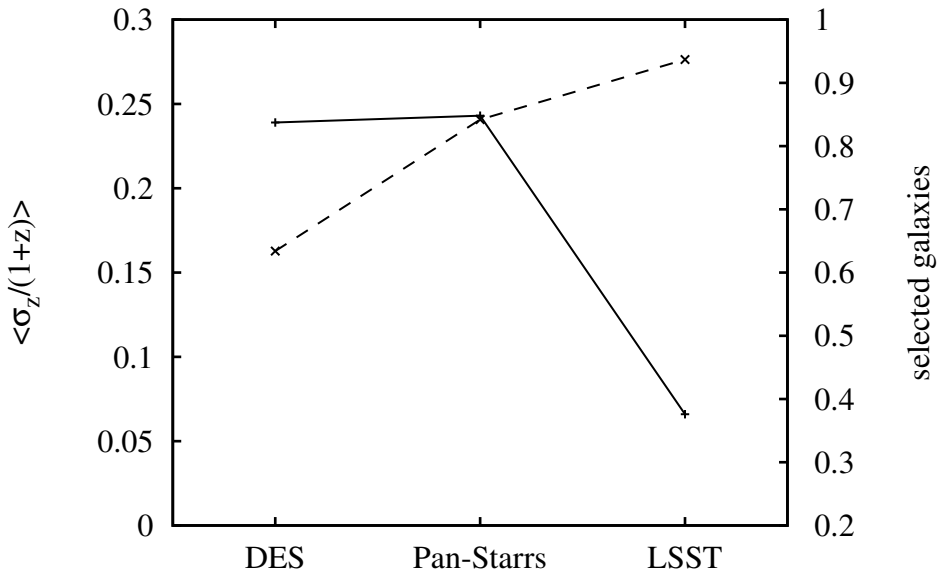

Figure 5.15: Quality of the photometric redshifts obtained with the different ground-based surveys. The solid line represents $\langle \frac{\sigma_z}{1+z} \rangle$, while the dashed line represents the fraction of galaxies with observed RIZ magnitude < 24.5 that are available after a cut is performed to reduce the redshift standard deviation below 0.05.

struments. We see that, even with the most severe selection, the redshift standard deviation remains above 0.03 in all three cases. This is due to very unfortunate overlapping of galaxies, that were detected as single objects by our routines. The redshift deduced by the combined photometry of the two galaxies is then compared to the one of the most brilliant object in the RIZ band, that is sometimes totally different. This kind of problem depends very mildly on the number of bands and on the quality of the photometry and has been only partially solved running a detection on the H-band image, in addition to the one on the RIZ image (See Sect. 5.1.3).

The trade-off between redshift precision and galaxy fraction is also strongly connected to the effectiveness of the procedure of removing galaxies that are likely to have a wrong redshift. We used a selection via ODDS, while other authors preferred using the 1-σ error quoted by the photometric redshift code (Abdalla et al., 2008), or investigating the presence of secondary peaks (Bordoloi et al., 2010). We tried these other methods, but the results were in general poorer that the ones we obtained with ODDS. If we were able to make an ideal selection, knowing perfectly what are the most wrong redshifts in the sample, we would need to remove $\sim 1\%$ of the galaxies with the LSST data, and $\sim 5\%$ with the DES and PanStarrs ones, to get to the required precision.

Our simulations are of course a simplified version of what is done to measure redshifts from multi-band real images, nonetheless they are more realistic than the simple creation of dirty catalogues. Thus, is not surprising that our results are in general worse than what is obtained by previous works done with that method (Abdalla et al., 2008; Bordoloi et al., 2010). For example, Bordoloi et al. (2010) find, for a DES-like survey, a standard deviation of the selected sample equal to 0.064 (so in fact higher than our threshold) after removing 12 % of the galaxies. As can be seen in Fig. 5.16, we instead would need to eliminate ~ 15 % of the sample to get to the same precision. To better understand this difference, we made some tests using simulated catalogues created without going through image simulation. We calculated magnitudes in the necessary bands from the theoretical SEDs of the galaxies in our sample, and we cut the resulting catalogue at RIZ $<$ 24.5. Then, we added Gaussian noise according to the observing depth for the Euclid-like survey and for the three realistic ground-based surveys. The

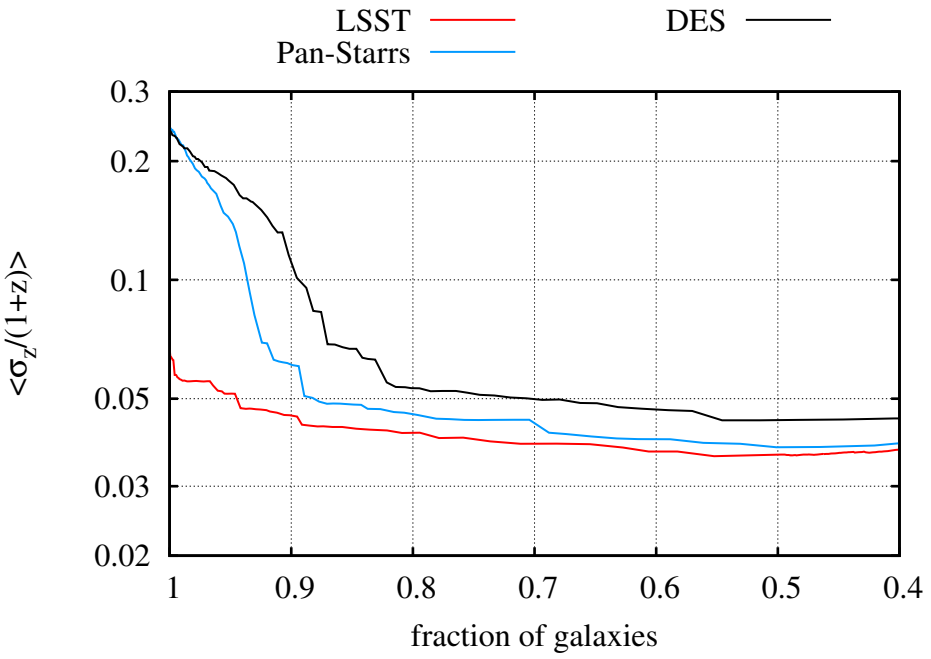

Figure 5.16: Redshift standard deviation as a function of the fraction of galaxies kept in the sample via a ODDS selection for the three simulated surveys.

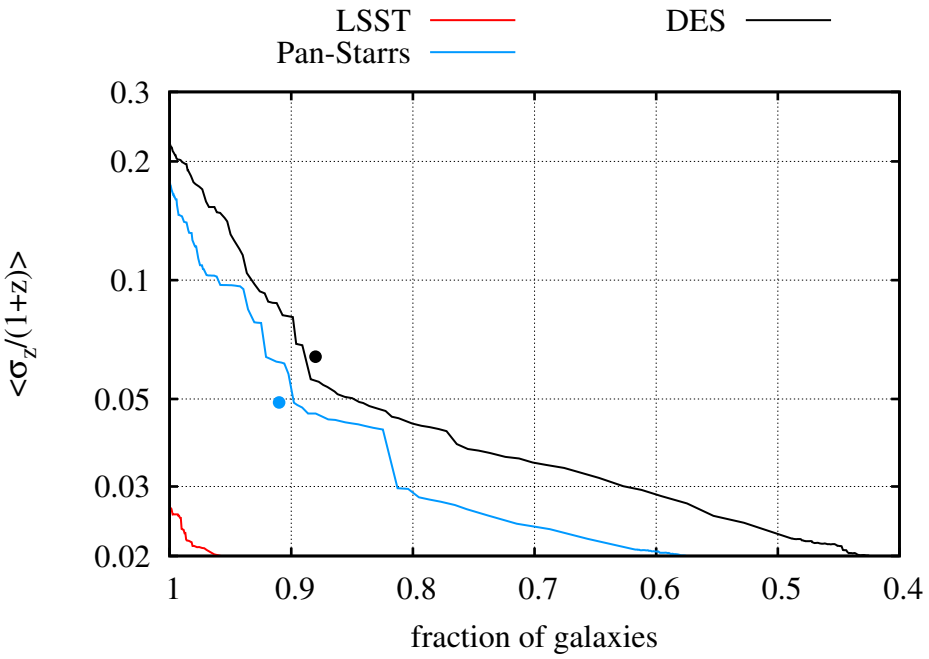

Figure 5.17: As Fig. 5.16, but from mock catalogues obtained without going through image simulation. The points represent the results obtained by Bordoloi et al. (2010).

resulting multi-band catalogues have been submitted to BPZ and analysed in the same way as those obtained from Colorpro on the simulated images. In Fig. 5.17 we show the results, namely the redshift standard deviation as a function of the fraction of galaxies kept in the catalogue. The results we obtain in this way are significantly better than the ones obtained with the complete simulation (Fig. 5.16), and are in close agreement with Bordoloi et al. (2010).

As can be seen from Figs. 5.16 and 5.17, the improvement we obtain with the simpler set of simulations is relevant especially in the cleaning procedure, that becomes more effective. This may depend on the fact that in this case the errors quoted in the photometry catalogue are by construction the real uncertainties on the magnitudes, while the procedure of multiband photometry when dealing with images can bring hidden errors, such as those due to the matching of images with different quality. As the errors in the photometry translate in the uncertainty in the redshift determination, the

value of the ODDS parameter is certainly more trustable in the simpler case and cleaning the catalogue from possible catastrophic outliers is thus more straightforward.

The work of Abdalla et al. (2008) differs in that they use the neural-network code ANNz to compute the redshifts. This may explain some differences on the importance of different bands between this work and theirs. In fact, they quote a σ_z (note: without dividing by $1+z$) of 0.307 for a DES survey, 0.274 for Pan-Starrs, 0.197 for LSST. We get 0.398 (0.359 with the 'simple' simulations) in the DES case, 0.400 (0.311) for Pan-Starrs, 0.146 (0.131) for LSST. Thus, they perform slightly better in the *griz* cases, while we do better when the u-band is available. This reflects also in the fact that they are able to get a $\sigma_z = 0.113$, below the LSST case, for an 'ideal' optical survey, while we do not obtain such a result, as shown in Section 5.1.7. This is reasonably due to the completely different techniques adopted. Although in general they get comparable results, it is possible that in this case we are showing strengths and weaknesses on both sides. Of course a firm statement on this topic can be done only submitting the same photometric catalogue to different photo-z algorithms.

5.1.9 Summary and conclusions

We investigated the quality of photometric redshifts that can be obtained for the galaxies observed by future space-based missions targeted to measure cosmic shear. In order to create a realistic setup for our work, we chose to simulate a Euclid-like mission, with imaging in an optical band and multi-band photometry in the near infrared (Laureijs et al., 2009). Fulfilling the requirement on photometric redshift precision is a crucial step to reduce the error in cosmic shear measurement, and it is important as well in other scientific subjects that can be studied with the same data (e.g. galaxy clusters, galaxy evolution). To this end, the space-based survey will need the collaboration of ground-based multi-band observations to characterise the SED of the galaxies in the optical range. We studied this topic using complete simulations, from realistic images of a patch of the sky, obtained with Skylens (Meneghetti et al., 2008), through multi-band photometry, to the photometric redshift estimation itself, performed with a template-fitting method (Benitez, 2000). This allowed us to consider many sources of uncertainty in

the process of determining a galaxy photo-z. Our main aim was to determine what fraction of the galaxy sample observed by a Euclid-like survey must be discarded to achieve $\langle \frac{\sigma_z}{1+z} \rangle < 0.05$, a precision considered suitable for cosmic shear precision measurements.

We first studied the impact of different parameters (depth, seeing, bands) on the quality of the redshift estimation. We found that:

- The number of good photo-zs increases as expected with the exposure time, as more and more galaxies acquire a good photometry that enables to assign to them a certain redshift. The improvement is not so relevant if we instead consider the mean precision of the whole sample, as some pathological galaxies (affected by light of neighbouring objects, or by severe colour-redshift degeneracies) pollute the catalogue even with the deepest survey we simulated.

- Seeing is potentially as important as depth, as blurring of the images is a great source of noise in the photometry, especially for FWHM > 1 arcsec. Increasing the seeing from 0.8 to 1.2 arcsec, we need to remove twice the number of galaxies to reach the same precision in photo-z estimation.

- Having u-band observations seems very important to improve the goodness of photo-zs. Even a relatively shallow survey with a 10-σ magnitude limit equal to 23.5 has a great impact on the redshift standard deviation of the total sample. Ultraviolet light is particularly relevant for galaxies at redshift below 0.5 and above 2.5 and resolves a lot of magnitude-colour degeneracies between these two ranges. The impact is not so strong when dealing with the fraction of available galaxies: in this case the improvement is around 10 % of the total sample in comparison with a corresponding *griz* survey.

Then, we moved to some more specific tests, aimed to study the results that would be obtained with some of the ground-based surveys that are currently considered as possible Euclid counterparts. We focused on three projects: DES, Pan-Starrs and LSST. DES and Pan-Starrs suffer the lack of u-band observations, and to clear the catalogue from the catastrophic failures and get to the required precision we need to eliminate respectively ~ 35 % and

~ 15 % of the galaxies of the sample. LSST performs much better than the other two, and seems to be the best, even if the farthest in time, counterpart to a space-based cosmic-shear mission.

With the aim of comparing our results with previous works, we made also some simpler simulations, creating catalogues with known photometric uncertainties without making mock observations. In this way, as expected, the results improve significantly, especially in the cleaning procedure. This means that the depth of the survey alone does not include all the information about the noise of the observation. Other uncertainties are added because of the proximity of galaxies, that pollute each other's photometry, and to the necessity of matching images obtained with different observing conditions and telescopes. A possible solution to the problems caused by the proximity of galaxies would be to use some automatic algorithm to eliminate a priori objects that appear to be very close to each other or even completely blended (i.e., detected as a single object). This points towards the need of even more realistic simulations, where it will be possible to optimize the procedure of object detection and multi-band photometry extraction from the images, a topic we did not tackle in this work. For example Kuijken et al. (2008) proposed a more complex approach, that makes use of shapelets decomposition to model both the source and the point-spread function.

We want to stress here that we used a photometric redshift code that has performed well in other applications to real and simulated data, but we did not select it as the best in the treatment of the specific data set we applied it to, nor we optimised it in this view. Our sample cleaning method could maybe be improved too, e.g. tuning the integration limits in the ODDS calculation, or considering secondary peaks, as done by Bordoloi et al. (2010), or the 1-σ error quoted by the photometric redshift code, as done by Abdalla et al. (2008). In general, our results should not be considered as optimal forecasts for a Euclid-like experiment, but more as a quantification of strengths and challenges for this kind of mission. In particular, the difference between results obtained with noisy catalogues (Fig. 5.17) and with image simulations (Fig. 5.16) hints at the importance of multi-band photometry extraction and *a-priori* object selection for this kind of measurement. Further testing and development of the algorithms is needed to understand what is the optimal way to extract photometric redshifts from this kind of observations

and to optimise the methods for each single case.

Overall, with the cautions we underlined above, our work show that it is indeed possible to obtain precise ($\langle\frac{\sigma_z}{1+z}\rangle < 0.05$) and robust photometric redshifts for more than 90 % of the galaxies with RIZ < 24.5, in case u-band observations are available. In the case of *griz* ground-based observations, on the other hand, to get to the required precision we need to eliminate at least 15 % of the sample (See Fig. 5.16), depending on the depth of the survey.

5.2 Cluster counts

5.2.1 Creation of mock catalogues

We now estimate the amount of galaxy clusters detectable by the Euclid survey via photometric observations. We perform this task by creating realistic mock catalogues according to the characteristics of the survey and by submitting them to the optimal filter described in Sect. 4.3. In principle, the work should start from simulations of observations like the ones performed in the previous Section, with the inclusion of realistic galaxy clusters on top of the distribution of field galaxies. We decided instead as a first approach to directly create a photometric multi-band catalogue, with the inclusion of photometric redshift measurements. In the creation of the catalogue, we make use of the knowledge we gained from the work described in the previous Section about the photometric redshift accuracy. In particular, we create a 5×5 deg^2 mock light-cone which is made of two components: a background galaxy field and a realistic population of galaxy clusters.

The distribution of field galaxies is created starting from the COSMOS photo-z catalogue (Ilbert et al., 2009), which we already introduced in Sect. 4.4. Magnitudes in the Euclid bands are assigned to each galaxy in the catalogue according to their spectral-energy distribution, which has been selected in the process of redshift estimation. Given that we do not need high-resolution morphology for this work, we preferred this catalogue with respect to the HUDF one (Coe et al., 2006) because it covers a larger area and allows us to span a wider range of galaxy properties. Only galaxies with $\mathrm{mag_H} < 24$ and $\mathrm{mag_Y} < 24$ are considered, following the expected depth of Euclid infrared observations (see Sect. 2.4), in good agreement with tests we made on

simulated images. Galaxy positions are assigned randomly inside the field of view. The total field catalogue is made of around 4×10^7 galaxies.

The creation of the cluster galaxy component starts from the compilation of a cluster catalogue, following the theoretical mass function, which we described in Sect. 1.2.4. In this work we considered a standard ΛCDM model with $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$ and $\sigma_8 = 0.8$. The positions of the clusters are assigned randomly inside the field of view. We focused on structures with redshift $0 < z < 2$ and mass above $10^{13.5}$ $M_\odot$, for a total of 2376 objects. From this cluster catalogue, two galaxy catalogues are created. The first one, which we call 'ideal', follows the model described in Sect. 4.3.3. To verify the dependence of the results of cluster detection on the characteristics of the cluster population, we then create a second 'scattered' cluster galaxy catalogue. In this case, we introduce a 20 % Gaussian scatter in three parameters that describe the galaxy population: the cluster richness, the red fraction and the $L_\star$ parameter of the luminosity function. Both the 'ideal' and the 'scattered' catalogues are made of around 2×10^6 galaxies. We then add each cluster galaxy catalogue to the field catalogue, obtaining two complete realistic catalogues.

A sensible step in the catalogue creation is the assignment of photometric redshifts to the galaxies. We perform this task making use of the knowledge we gained with the work presented in the previous Section. In particular, we use the catalogue obtained from multi-band realistic observations with Euclid and LSST. For each galaxy in the mock catalogue, we choose in the Euclid+LSST catalogue the most similar object in terms of photometry and redshift, and we use its error and the quoted 1-σ uncertainty in the redshift measurement. In this way, we assign to each galaxy a realistic redshift measurement without the need to go again through the whole procedure described in the previous Section.

5.2.2 Results

In the analysis, we choose to consider the infrared bands Y and H, which allow us to observe a larger number of objects with respect to the optical RIZ band. The model considered to create the filter has a mass M_{200} equal to 10^{14} $M_\odot$. We compute the estimates of Λ over a grid with an angular resolution equal to 0.5 R_{200} at redshift slices separated by $\Delta z = 0.05$, which is enough

considering the expected precision on photometric redshift measurements ($\sigma_z = 0.05 \times (1 + z)$). Peaks are extracted from the maps and a redshift estimate is assigned to cluster candidates following the procedure described in Sections (4.3) and (4.4). We then match cluster detections with the original cluster catalogue with a tolerance equal to $R_{200} \times \log(\Lambda^{0.42})$ in the angular position and 0.1 in the redshift. The factor proportional to the richness Λ follows the relation between N_{200} and R_{200} and allows us to search for a counterpart in the input catalogue in an area that is reasonably linked to the effective size of the cluster. The redshift tolerance is larger than the one we adopted in Sect. 4.4 because of the different redshift precision of the two catalogues.

In the analysis of the 'ideal' catalogue, we obtain 898 cluster detections with this procedure. With the linking lengths described above, we get 853 associations between input clusters and detections. As there are both single detections associated with more than an input cluster and vice versa, from this value we cannot get any conclusion about purity and completeness of our sample. We first evaluate the purity of our sample by calculating the fraction of detections above a given threshold which have at least a counterpart in the cluster catalogue. The results are shown in Fig. 5.18. We see that our detected sample has a purity around 90% for S/N > 4 and it is very close to be perfectly pure for S/N > 5. This last result confirms what we got from the analysis of the COSMOS field (see Fig. 4.15).

We now show the amount of detected clusters as a function of their mass, considering both the S/N $= 3$ and the S/N $= 4$ thresholds. The results are shown in Fig. 5.19. Considering the safer threshold (S/N $= 4$) for detections, we get a $> 80\%$ completeness above $\log M/M_\odot = 14.1$. In this simulation, this corresponds to ~ 3 objects per square degree. This value confirms the expected total cluster counts in the Euclid survey, which is 60000 (see Laureijs et al., 2011). We note here that these results are slightly better than those we obtained naively applying to all the galaxies a redshift error equal to the requirement, $\sigma_z = 0.05 \times (1 + z)$, as we show in Fig. 5.20. This depends on the fact that many bright galaxies have a redshift estimate which is more precise than the requirement. Given that the most brilliant galaxies are those that receive larger weights from our algorithm, and thus that drive the cluster detection, the high quality of their photometric redshifts helps to pinpoint big

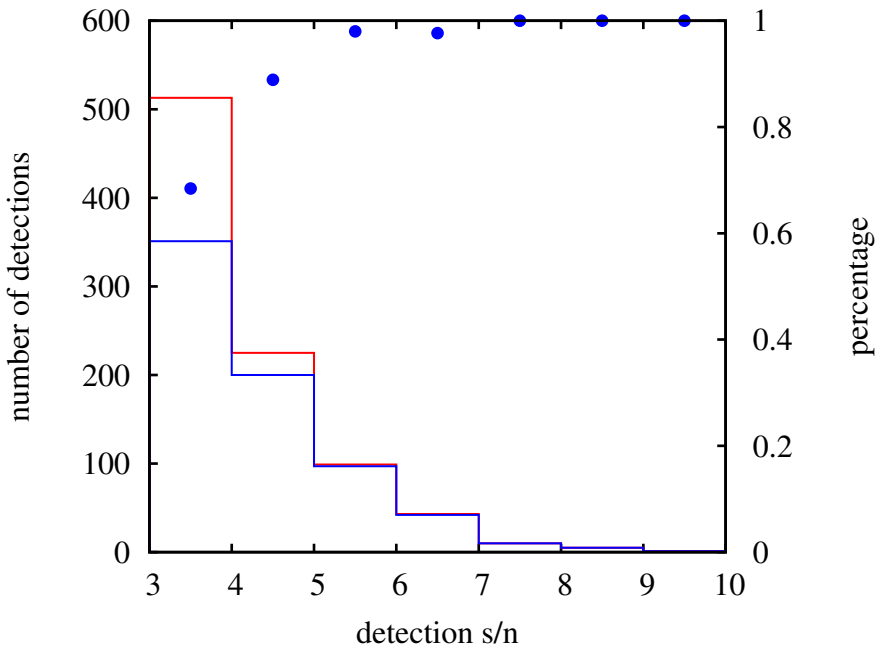

Figure 5.18: Amount of detections at a given significance which have at least a counterpart in the cluster input catalogue. The red histogram represents the total amount of detections, the blue histogram the ones with at least a counterpart, the blue points the fraction of confirmed detections.

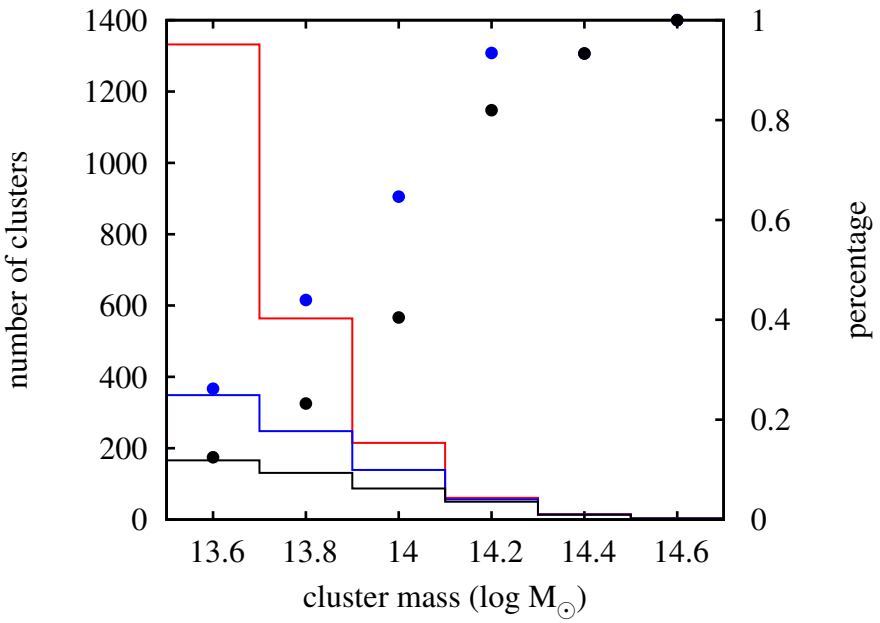

Figure 5.19: Mass distribution of the total cluster sample (red histogram), of the ones detected at S/N $>$ 3 (blue histogram) and of the ones detected at S/N $>$ 4 (black histogram). The blue points represents the fraction of clusters detected at S/N $>$ 3, the black points the fraction of clusters detected at S/N $>$ 4 (scale on the right).

structures in the analysis. This compensates for the fact that many dimmer galaxies have redshift uncertainties bigger than the requirement, and suffer from catastrophic failures, as shown in the previous Chapter.

To extract information on the cosmological parameters from cluster counts, high redshift ($z > 1$) detections are very important. Thus we measure the abundance of cluster detections as a function of redshift, as shown in Fig. 5.21. We see that the fraction of detected clusters declines rapidly with redshift above $z = 1$. We must use detections at S/N $>$ 3 to have a large enough sample of detected clusters, none of which has a redshift larger than 1.6. The amount of detected clusters above $z = 1$ is expected to be around 10000 by Laureijs et al. (2011), with an angular density equal to ~ 0.7 objects per deg^2. We obtain around 0.5/deg^2 counts only if we consider detections at S/N $>$ 3.

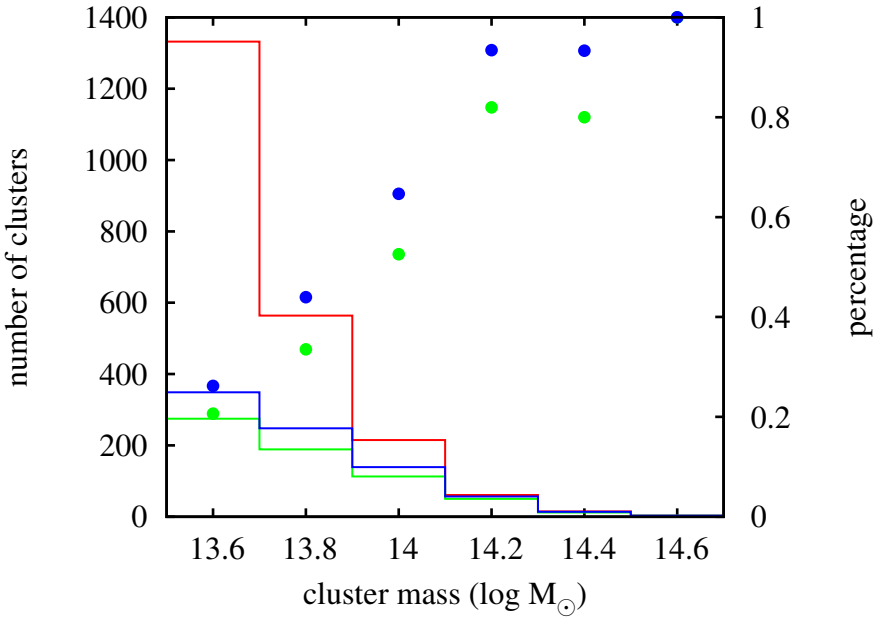

Figure 5.20: Mass distribution of the total cluster sample (red histogram), of the ones detected in the catalogue with realistic photo-z errors (blue histogram) and of the ones detected in the catalogue with photo-z errors that match the Euclid requirements (green histogram). The blue points represents the fraction of detected clusters in the former catalogue, the green points in the latter (scale on the right). The detection threshold is assumed to be S/N = 3 in both cases.

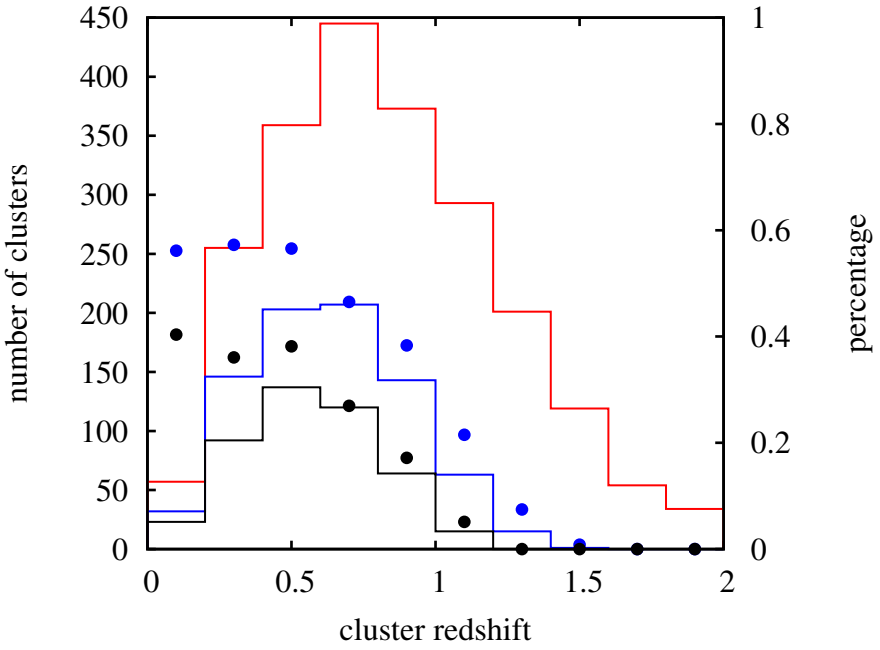

Figure 5.21: Redshift distribution of the total cluster sample (red histogram), of the ones detected at S/N $>$ 3 (blue histogram) and of the ones detected at S/N $>$ 4 (black histogram). The blue points represents the fraction of clusters detected at S/N $>$ 3, the black points the fraction of clusters detected at S/N $>$ 4 (scale on the right).

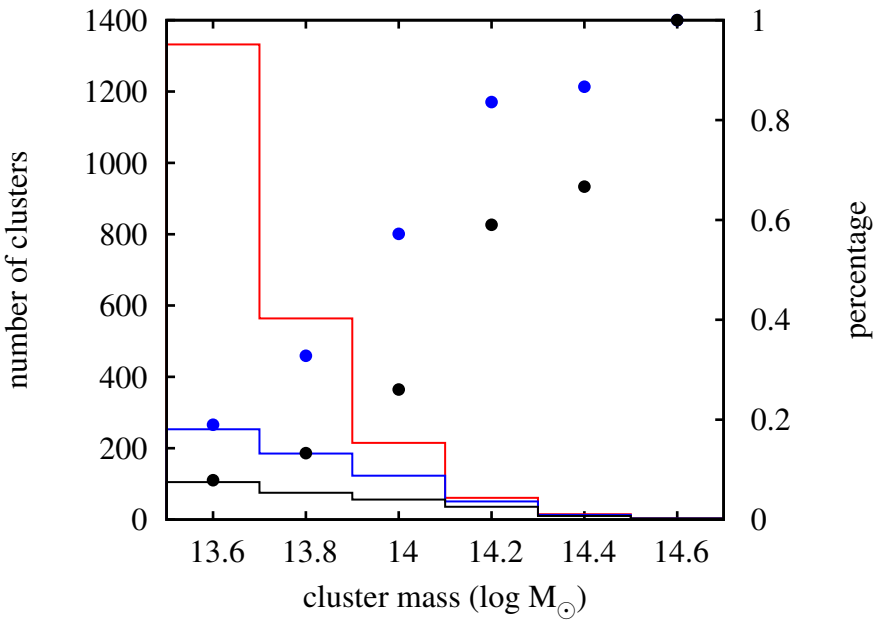

Figure 5.22: As Fig. 5.19, but for the 'scattered' galaxy cluster catalogue.

We now consider the analysis of the 'scattered' catalogue, that contains clusters with a galaxy population that deviates from our model. The completeness as a function of mass and redshift is shown in Figs. 5.22 and 5.23, respectively, while the purity of the sample is shown in Fig. 5.24. The total number of matches between the detections and the input catalogue lowers from 853 to 653 changing the cluster galaxy distribution. The filter is optimised to detect clusters which follow the model we built starting from observed data. When we modify the parameters of the clusters present in the catalogue, the signal-to-noise of the detection of each cluster gets lower and some objects go under the S/N = 3 threshold. As a result of this, comparing Figs. 5.19 and 5.22 we see that the completeness as a function of mass of detections at S/N $>$ 3 in the 'scattered' catalogue is very similar to the one of detections at S/N $>$ 4 in the original catalogue. Unfortunately, this results in an increased impurity of the sample, as can be seen in Fig. 5.24, because detections with a significance between S/N = 3 and S/N = 4 contain a $\sim$ 30 % rate of false positives. The redshift distribution of the detections is very similar to the 'ideal' case, as can be seen in Fig. 5.23.

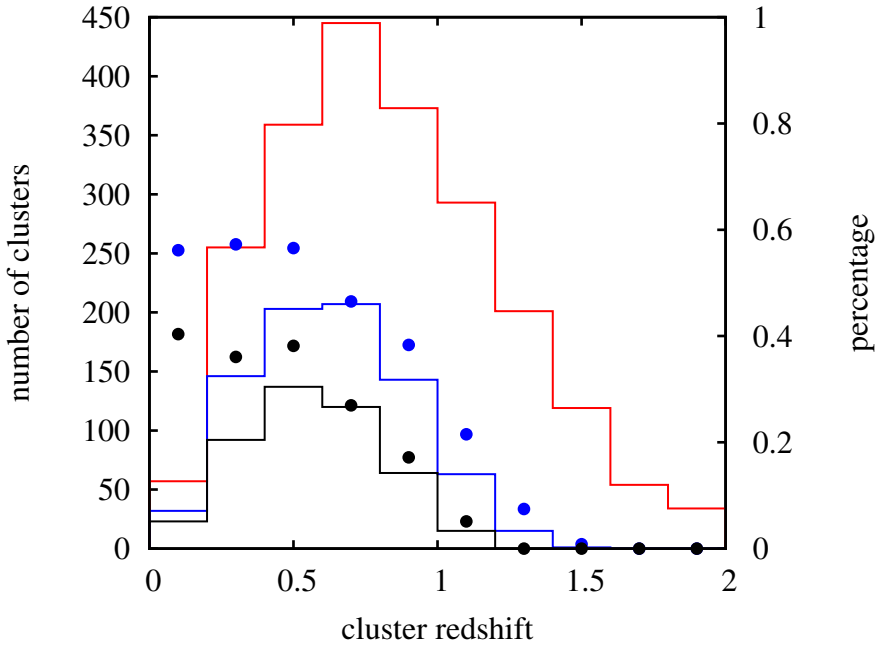

Figure 5.23: As Fig. 5.21, but for the 'scattered' galaxy cluster catalogue.

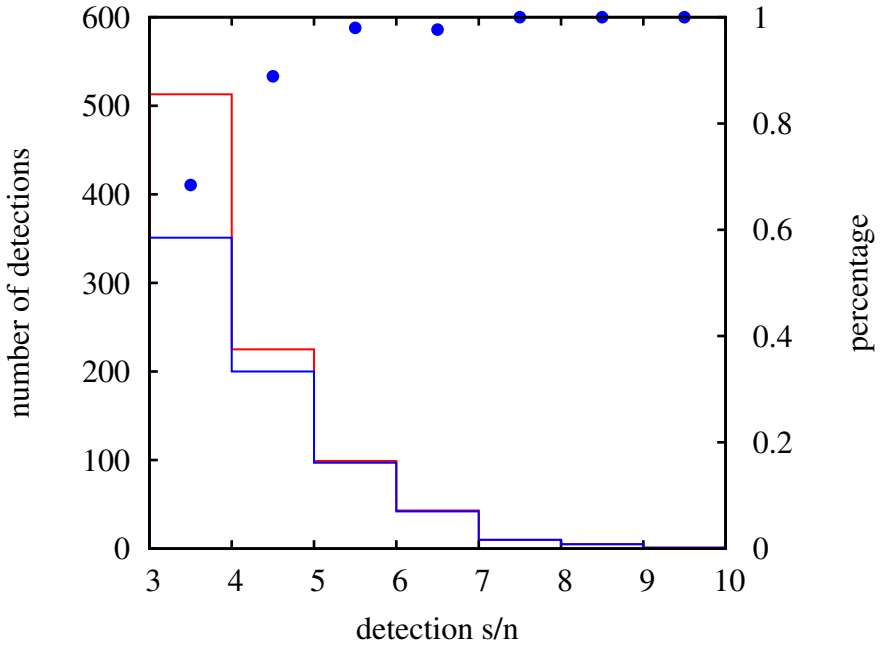

Figure 5.24: As Fig. 5.18, but for the 'scattered' galaxy cluster catalogue.

5.2.3 Tests on cosmological simulations

We now test our algorithm on a catalogue extracted from a cosmological simulation. We make use of the 'Magneticum Pathfinder' simulation run with the Gadget-3 code (Springel, 2005; Dolag & Stasyszyn, 2009), which is a tree-smoothed particle hydrodynamics code that fully conserves entropy during the evolution of the gas component, taking into account radiative cooling, heating by an ultraviolet meta-galactic background to emulate the reionization era and a subresolution scheme to treat star formation, supernova feedback and galactic winds. The data are taken from the simulation with a box-size equal to 128 Mpc/h, and a particle mass equal to 3.6×10^7 $M_\odot/h$ for dark matter and 7.3×10^6 $M_\odot/h$ for baryons. For the star formation recipe it is assumed that each star particle of the simulation is a single stellar population where the relative number of stars with different masses is obtained by means of the initial mass function, which is that of Salpeter (1955) normalized in the 0.1-100 $M_\odot$ mass range. The luminosities of the simulated galaxies are computed using the stellar population synthesis model of Bruzual & Charlot (2003) in different spectral bands, summing up the contributes of each stellar population.

The resulting light-cone has a side equal to 3 degrees and a limiting redshift $z = 1$. From the resulting galaxy catalogue, we select the objects within the Euclid limiting depth in the near-infrared bands and we change their redshifts according to the observational uncertainties as described in Section 5.2.1. We then submit the resulting catalogue to our algorithm and we perform the search as usual in redshift slices with a distance Δz =0.05r. We then match the resulting detections to the catalogue of structures extracted directly from the simulation boxes, which is limited to $M = 10^{14}$ $M_\odot$. The resulting completeness as a function of mass is shown in Fig. 5.25.

Although the sample is quite small, we confirm the ability to detect structures with mass M above $10^{14.2}M_\odot$.

5.2.4 Summary and future prospects

In this Section we tested the capability of detecting clusters through Euclid photometric data. For this purpose, we applied the optimal filter introduced in Sect. 4.3 on mock catalogues obtained from other observations and from

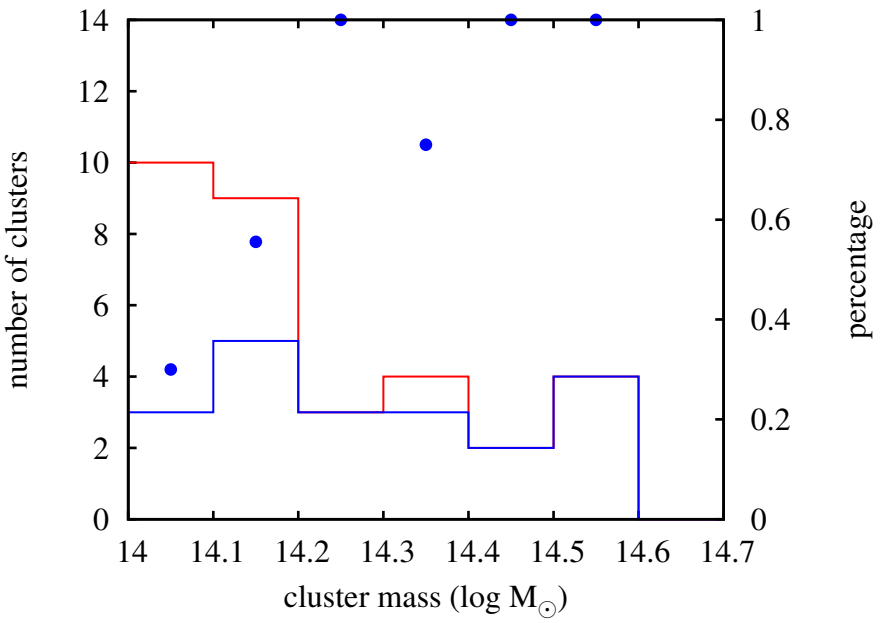

Figure 5.25: Completeness as a function of mass for the test done on catalogues extracted from a cosmological simulation. The symbols are the same used in Fig. 5.19.

cosmological simulations. The catalogues were cut at the expected Euclid limiting depth, and photometric redshifts were assigned to the galaxies according to the precision we calculated in Sect. 5.1. In the case where the clusters were drawn from the model used in the filter, we get a $> 80\%$ completeness above $\log M/M_{\odot} = 14.1$, considering safe detections at S/N = 4. When a scatter in the cluster properties is allowed, the signal-to-noise of the detections slightly worsens, and we need to use less safe detections at S/N = 3 to reach the same completeness. The application of the filter on catalogues extracted from cosmological simulations shows a very similar trend, although the catalogue is too limited to draw quantitative conclusions. Nevertheless, this test was important to assess the ability of our algorithm to detect structures in a more realistic scenario, where we had no control on the properties of the input galaxy clusters. A summary of the 70 % detection threshold as a function of redshift is shown in Fig. 5.26.

In the future we wish to extend this work by making use of a light-cone with a larger angular size and a higher redshift limit and by performing end-to-end simulations, that start from mock observations. The larger size is required to improve the statistics and to include a more varied population of objects in the data. The higher redshift limit will allow us to test the ability to detect clusters in a very important range, where clusters counts are very sensible to the cosmological parameters. With the usage of mock observations we can introduce more realistic sources of noise: for example, the density of the galaxy population in clusters can produce bigger uncertainties in the estimation of photometry and redshift. Moreover, we can introduce the lensing effects on galaxy observed shapes and then run a combined search using information coming from both weak lensing data and galaxy distribution, as we did on the COSMOS field. We will then be able to assess in a more robust way the capability to detect galaxy clusters in Euclid data.

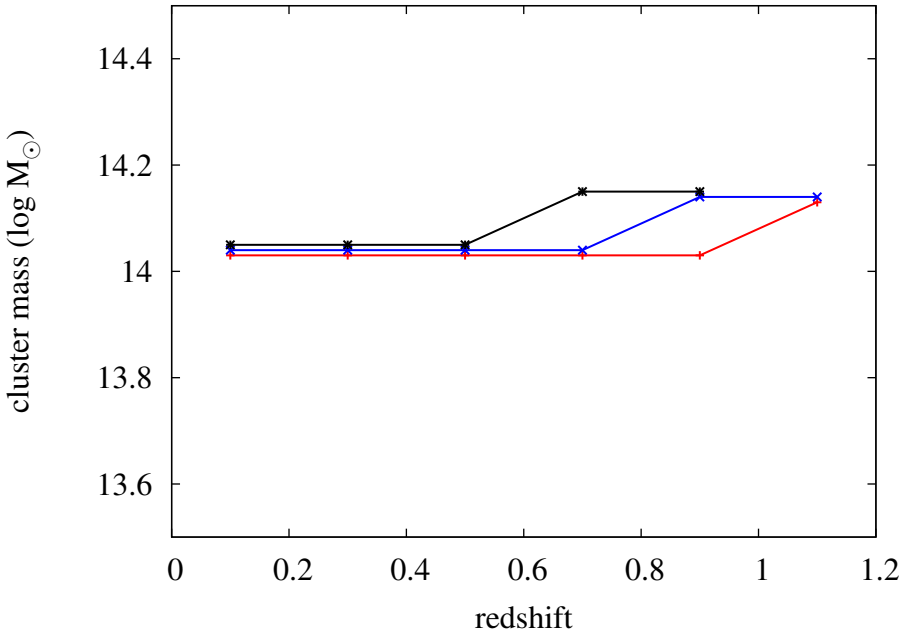

Figure 5.26: Mass threshold for a 70 % detection rate in cluster detection. The red line refers to the analysis of the 'ideal' catalogue, while the blue line refers to the 'scattered' catalogue. The black line refers to the analysis of the catalogue extracted from the Magneticum Pathfinder simulation.

.

Conclusions

In this Thesis we have presented our work on the production of mock simulations of observations and on their application to test the performances of future weak lensing surveys. Experiments such as the ESA-approved Euclid mission promise to constrain cosmological parameters with unprecedented accuracy through the analysis of the cosmic shear signal produced by the large-scale structure on the observed shape of galaxies. To fulfil this task, however, many systematics must be put under control and the analysis routines must be tested on synthetic realistic data. The creation of mock observations which contain many of the sources of noise present in real data allows us to assess the precision that can be obtained in the data analysis.

In Chapter 3, we have shown how simulations of observations can be created starting from observed galaxy catalogues and considering characteristics of the telescope and of the survey. We then applied this tool to verify the sources of errors and inconsistencies in the mass measurement of galaxy clusters from lensing data. We have shown that the alignment between the major axis of the cluster and the line of sight is an important source of error in lensing mass estimates, producing a 20 % scatter around the correct value. On the other hand, lensing does not suffer from the bias due to incorrect assumptions on the cluster gas physics, such as methods based on X-ray data.

In Chapter 4, we approached the problem of cluster detection from weak lensing and photometric data. We built an optimal filter for cluster detection that automatically selects the important features to distinguish a cluster population from the field one. After having tested it on mock data, we applied it to the COSMOS field together with its weak lensing counterpart. We presented a catalogue of 27 lensing-confirmed clusters, 11 of which do not have any previous detection in the literature. Comparing the sample of

detected clusters with previous catalogues obtained from X-ray observations, we verified that we were able to detect more than 50% of the clusters with an X-ray mass over 1.5×10^{13} $M_{\odot}$.

In Chapter 5, we applied the techniques presented in the previous Chapters to the Euclid mission. We first verified the precision in the photometric redshift measurements for galaxies detected by Euclid. Photometric redshifts are key data to perform lensing tomography and the Euclid mission will need ancillary data from ground-based facilities to properly sample the galaxy SEDs. We simulated observations with Euclid instruments and with fictional and real ground-based instruments. We verified the importance of good observing sites (with seeing < 1 arcsec) and of the presence of u-band observations to reduce the number of catastrophic failures. In particular, considering possible Euclid ground-based counterparts, we found that we must eliminate $\sim 15\%$ of the sample to reach the required precision in case LSST data are used, while the discarded fraction rises to 35% in case Pan-Starrs or DES data are considered.

We then applied our cluster finder algorithm to Euclid mock data to assess the amount of galaxy clusters that can be observed. We used both catalogues created from observations with *ad-hoc* clusters on top of the field population, and catalogues extracted from cosmological simulations. In both cases, we verified that is possible to get an almost complete sample for masses above $\log M_{\odot} = 14.2$ and $z < 1$, with a low ($\sim 20\%$) rate of false detections. Above redshift $z = 1$, the cluster detection becomes more difficult, but the data were not sufficient to draw conclusions on the limiting mass. In total, the number of clusters detected from Euclid photometric data should be close to the expected value of 60000.

Bibliography

Abell P.A., Allison J., Anderson S.F., et al., 2009, arXiv preprint, astro-ph/0912.0201

Abbott T., Aldering G., Annis J., et al., 2005, arXiv preprint, astro-ph/0510346

Abdalla F.B., Amara A., Capak P., Cypriano E.S., Lahav O., Rhodes J., 2008, MNRAS, 387, 969

Abdalla F.B., Banerji M., Lahav O., Rashkov V., 2011, MNRAS, 417, 1891

Aldering G., Althouse W., Amanullah R., et al., 2004, PASP, submitted; preprint astro-ph/0405232

Amara A., Refregier A., 2007, MNRAS, 381, 1018

Balogh M.L., Baldry I.K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, ApJ, 615,101

Bartelmann M., 1996, A&A, 313, 697

Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291

Beckwith S.V.W., Stiavelli M., Koekemoer A.M., et al., 2006, AJ, 132, 1729

Bellagamba F., Maturi M., Hamana T., Meneghetti M., Miyazaki S., Moscardini L., 2011, MNRAS, 413, 1145

Bellagamba F., Meneghetti M., Moscardini L., Bolzonella M., 2012, MNRAS, in press, arXiv preprint, astro-ph/1201.3590

Benitez N., 2000, ApJ, 536, 571

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bordoloi R., Lilly S.J., Amara A., 2010, MNRAS, 406, 881

Bolzonella M., Miralles J.-M., Pello R., 2000, A&A, 363, 476

Bridle S., Balan S.T., Bethge M., et al., 2010, MNRAS, 405, 2044

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Carroll S.M., Press W.H., Turner E.L., 1992, ARA&A, 30, 499

Clowe D., Luppino G.A., Kaiser N., Henry J.P., Gioia I.M., 1998, ApJ, 497, L61

Coe D., Benitez N., Sanchez S.F., Jee M., Bouwens R., Ford H., 2006, ApJ, 132, 926

Coleman G.D., Wu C.-C., Weedman D.M., 1980, ApJS, 43, 393

Collister A.A., Lahav O., 2004, PASP, 116, 345

Dahlen T., Mobasher B., Dickinson M., et al., 2010, ApJ, 724, 425

Davis M., Peebles P.J.E., 1983, ApJ, 267, 465

Dietrich J.P., Erben T., Lamer G., Schneider P., Schwope A., Hartlap J., Maturi M., 2007, A&A, 470, 821

Dolag K., Stasyszyn E., 2009, MNRAS, 298, 1697

Dong F., Pierpaoli E., Gunn J.E., Wechsler R.H., 2008, ApJ, 676, 868

Els S.G., Schöck M., Bustos E., et al., 2009, PASP, 121, 922

Erben T., Hildebrandt H., Lerchster M., 2009, A&A, 493, 1197

Fahlman G., Kaiser N., Squires G., Woods D., 1994, ApJ, 437, 56

Feldmann R., Carollo C.M., Porciani C., 2006, MNRAS, 372, 565

Finoguenov A., Guzzo L., Hasinger G., et al., 2007, ApJS, 172, 182

Frieman J., Turner M.S., Huterer D., 2008, ARA&A, 46, 385

Gilbank D.G., Bower R.G., Castander F.J., Ziegler B.L., 2004, MNRAS, 348, 551

Gladders M., Yee H.K.C., 2000, AJ, 120, 2148

Gladders M., Yee H K.C., 2005, AJ, 157, 1

Grazian A., Fontana A., De Santis C., Gallozzi S., Giallongo E., Di Pangrazio F., 2004, PASP, 116, 750

Hamana T., Miyazaki S., Kashikawa N., Ellis R.E., Massey R.J., Refregier A., Taylor A., 2009, PASJ, 61, 833

Hansen S.M., McKay T.A., Wechsler R.H., Annis J., Sheldon E.S., Kimball A., 2005, ApJ, 633, 122

Hansen S.M., Sheldon E.S., Wechsler R.H., Koester B.P., 2009, ApJ, 699, 1333

Hao J., Koester B.P., Mckay T.A., et al., 2009, ApJ, 702, 745

Hennawi J.F., Spergel D.N., 2005, ApJ, 624, 59

Hildebrandt H., Wolf C., Benitez N., 2008, A&A, 480, 703

Hildebrandt H., van Waerbeke L., Erben T., 2009, A&A, 507, 683

Hoekstra H., Franx M, Kuijken K., Squires G., 1998, ApJ, 504, 636

Hu W., 1999, ApJ, 522, L21

Ilbert O., Capak P., Salvato M., et al., 2009, ApJ, 690, 1236

Kaiser N., Squires G., Broadhurst T., 1995, ApJ, 449, 460

Kaiser N., Aussel H., Burke B., et al., 2002, SPIE, 4836, 154

Kasliwal M., Massey R., Ellis R., Miyazaki S., Rhodes J., 2007, AAS, 38, 927

Kepner J., Fan X., Bahcall N., Gunn J., Lupton R., Xu G., 1999, ApJ, 517, 78

Kitching T.D., Miller L., Heymans C.E., van Waerbeke L., Heavens A.F., 2008, MNRAS, 390, 149

Kitching T., Balan S., Bernstein G., et al., 2010, arXiv pre-print, astro-ph/1009.0779

Kinney A.L., Calzetti D., Bohlin R.C., McQuade K., Storchi-Bergmann T., Schmitt H.R. 1996, ApJ, 467, 38

Koester B.P., McKay T.A., Annis J., et al., 2007, ApJ, 660, 221

Kuijken K., 2008, A&A, 482, 1053

Laureijs R., et al., 2009, arXiv preprint, astro-ph/0912.0914

Laureijs R., et al., 2011, arXiv preprint, astro-ph/1110.3193

Leinert C., Bowyer S., Haikala L.K., et al., 1998, A&AS, 127, 1

Li H.I., Yee H.K.C., 2008, AJ, 135, 809

Loh Y.-S., Ellingson E., Yee H.K.C., Gilbank D.G., Gladders M.D., Barrientos L.F., 2008, ApJ, 680, 214

Lu T., Gilbank D.G., Balogh M.L., Bognat A., 2009, MNRAS, 399, 1858

Luppino G.A., Kaiser N., 1997, ApJ, 475, 20

Magnier E., et al., 2007, ASPC, 364, 153

Massey R., Heymans C., BergÃĭ J., et al., 2007, MNRAS, 376, 13

Maturi M., Meneghetti M., Bartelmann M., Dolag K., Moscardini L., 2005, A&A, 442, 851

Maturi M., Schirmer M., Meneghetti M., Bartelmann M., Moscardini L., 2007, A&A, 462, 473

Mazure A., Adami C., Pierre M., 2007, A&A, 467, 49

Mei S., Holden B.P., Blakeslee J.P., et al., 2006, ApJ, 644, 759

Melchior P., Meneghetti M., Bartelmann M., 2007, A&A, 463, 1215

Menanteau F., Ford H.C., Motta V., Benitez N., Martel A.R., Blakeslee J.P., Infante L., 2006, AJ, 131, 208

Menanteau F., Hughes J.P., Jimenez R., 2009, ApJ, 698, 1221

Meneghetti M., Bartelmann M., Moscardini L., 2002, MNRAS, 340, 105

Meneghetti M., Melchior P., Grazian A., et al., 2008, A&A, 482, 403

Meneghetti M., Rasia E., Merten J., Bellagamba F., Ettori S., Mazzotta P., Dolag K., Marri S, 2010, A&A, 514, 93

Milkeraitis M., Van Waerbeke L., Heymans C., Hildebrandt H., Dietrich J.P., Erben T., 2009, MNRAS, 406, 673

Miyazaki S., Hamana T., Shimazaku K., et al., 2002, ApJ, 580, 97

Miyazaki S., Hamana T., Ellis R.S., Kashikawa N., Massey R.J., Taylor J., Refregier A., 2007, ApJ, 669, 714

Naraya R., Bartelmann M., 1996, astro-ph/9606001

Navarro, J.F., Frenk C.S., White S.D.M., 1997, ApJ, 490, 493

Olsen L.F., Benoist C., Cappi A., et al., 2007, A&A, 461, 81

Pace F., Maturi M., Meneghetti M., Bartelmann M., Moscardini L., Dolag K., 2007, A&A, 471, 731

Pace F., Maturi M., Bartelmann M., Cappelluti N., Dolag K., Meneghetti M., Moscardini L., 2008, A&A, 483, 389

Paulin-Henriksson S., Amara A., Voigt L., Refregier A., Bridle S.L., 2008, A&A, 484, 67

Perlmutter S., Aldering G., Goldhaber G., et al., 1999, ApJ, 517, 565

Popesso P., Böhringer H., Romaniello M., Voges W., 2005, A&A, 433, 415

Postman M., Lubin L.M., Gunn J.E., Oke J.B., Hoessel J.G., Schneider D.P., Christensen J., 1996, AJ, 111, 615

Press W.H., Schechter P., 1974, ApJ, 187, 425

Rafelski M., Wolfe A.M., Cooke J., Chen H.W., Armandroff T.E., Wirth G.D., 2009, ApJ, 703, 2033

Ramella M., Boschin W., Fadda D., Nonino M., 2001, A&A, 368, 776

Rasia E., Tormen G., Moscardini L., 2004, MNRAS, 351, 237

Refregier A., 2003, MNRAS, 338, 35

Riess A.G., Filippenko A.V., Challis P., et al., 1998, AJ, 116, 1009

Rowan-Robinson M., Babbedge T., Oliver S., et al., 2008, MNRAS, 386, 697

Salpeter E.E., 1955, ApJ, 121, 161

Schäfer B.M., Pfrommer C., Hell R.M., Bartelmann M., 2006, MNRAS, 370, 1713

Schechter P., 1976, ApJ, 203, 297

Sheth R.K., Tormen G., 1999, MNRAS, 308, 119

Schirmer M., Erben T., Schneider P., Wolf C., Meisenheimer K., 2004, A&A, 420, 75

Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, MNRAS, 296, 873

Schneider P., van Waerbeke L., Kilbinger M., Mellier Y., 2002, A&A, 396, 1

Schock M., Els S.G., Riddle R., et al., 2009, PASP, 121, 384

Scoville N., Aussel H., Brusa M., et al., 2007, ApJS, 172, 1

Scoville N., Aussel H., Benson A., et al., 2007, ApJS, 172, 150

Sheldon E.S., Johnston D.E., Masjedi M., 2009, ApJ, 703, 2232

Springel V., White S.D.M., Jenkins A., et al., 2005, Nat, 435, 629

Springel V., 2005, MNRAS, 364, 1105

Taniguchi Y., Scoville N., Murayama T., et al., 2007, ApJS, 172, 9

Thanjavur K., Willis J., Crampton D., 2009, ApJ, 706, 571

Tokovinin A., Travouillon T., 2006, MNRAS, 365, 1235

Turner M.S., White M.J., 1997, Phys. Rev. D., 56, 4439

Umetsu K., Medezinski E., Broadhurst T., Zitrin A., Okabe N., Hsieh B.C., Molnar S.M., 2010, ApJ, 714, 1470

Way M.J., Srivastava A.N., 2006, ApJ, 647, 102

Yee H.K.C., arXiv preprint, astro-ph/9809347

Ma allo stesso modo tu puoi trarne la conclusione opposta: che la vera mappa dell'universo sia la città d'Eudossia così com'è, una macchia che dilaga senza forma, con vie tutte a zig-zag, case che franano una sull'altra nel polverone, incendi, urla nel buio.
(Italo Calvino, Le città invisibili)

But you could, similarly, come to the opposite conclusion: that the true map of the universe is the city of Eudoxia, just as it is, a stain that spreads out shapelessly, with crooked streets, houses that crumble one upon the other amid clouds of dust, fires, screams in the darkness.
(Italo Calvino, The invisible cities)